LEVEL OF SERVICE CRITERIA OF ROADS IN URBAN INDIAN CONTEXT

ASHISH KUMAR PATNAIK

211CE3249

DEPARMENT OF CIVIL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA-769008

MAY.2013

LEVEL OF SERVICE CRITERIA OF ROADS IN URBAN INDIAN CONTEXT

A Thesis Submitted in Partial Fulfilment
Of the Requirements for the Award of the Degree of

Master of Technology

In

Civil Engineering

With Specialization in "Transportation Engineering"

By

ASHISH KUMAR PATNAIK

211CE3249

Under the Supervision of

Prof. Prasanta Kumar Bhuyan

DEPARMENT OF CIVIL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA-769008

MAY.2013

Department of Civil Engineering

National Institute of Technology, Rourkela

CERTIFICATE

This is to certify that the Thesis Report entitled **"LEVEL OF SERVICE CRITERIA OF ROADS IN URBAN INDIAN CONTEXT"** submitted by **Mr. ASHISH KUMAR PATNAIK** bearing roll no. **211CE3249** in partial fulfilment of the requirements for the award of Master of Technology in Civil Engineering with specialization in **"Transportation Engineering"** during session 2011-2013 at National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Place:

Date:

Prof. Prasanta Kumar Bhuyan
Assistant Professor
Dept. of Civil Engineering
National Institute of Technology
Rourkela-769008

Dedicated to

To My Loving Grand Father, My Parents, My Elder brother Jagannath, All Family Members & Teachers Whose efforts, sacrifice, patience, inspiration, and encouragement are helping me to move forward.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep sense of respect and gratitude towards my advisor and guide Prof. P.K. Bhuyan, who has been the guiding force behind this work. I am greatly indebted to him for his constant encouragement, invaluable advice and for propelling me further in every aspect of my academic life. His presence and optimism have provided an invaluable influence on my career and outlook for the future. I consider it my good fortune to have got an opportunity to work with such a wonderful person.

Next, I want to express my respects to our Director Prof. S. Sarangi, NIT Rourkela, HOD Civil Engineering Department Prof. N. Roy, Dean SRICCE Prof. M. Panda and Prof. U. Chattaraj for teaching me and also helping me how to learn. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I also extend my thanks to all faculty members and staff of the Department of Civil Engineering, National Institute of Technology, Rourkela who have encouraged me throughout the course of Master's Degree.

I would like to thank all my friends and especially my classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious. I have enjoyed their companionship so much during my stay at NIT, Rourkela.

I am especially indebted to my parents for their love, sacrifice, and support. They are my first teachers after I came to this world and have set great examples for me about how to live, study and work.

Ashish Kumar Patnaik

Date: Roll No: 211CE3249

Place: Dept. of Civil Engg.

NIT, Rourkela

ABSTRACT

India is a highly populated country; it has a second largest road network in the world. Owing to boastfully population, the congestion is growing at zip, zap, zoom speed as thousands of heterogeneous vehicles are added to the urban roads in India. The Level of Service (LOS) is not passable defined for heterogeneous traffic flow with different operational features. Delineating LOS is essentially a classification problem. The diligence of cluster analysis is the worthiest proficiency to solve such problem for which adaboost algorithm, Genetic programming, Maximum-Likelihood Method and Expectation-Maximization Method are used in this study. Five cluster validation parameters are utilized to examine the optimal no clusters. After acquiring optimal no of clusters, these four methods are implemented to the free flow speed (FFS) data to get ranges of different urban street classes. Again, these four clustering method are enforced on average travel speeds of street segments to specify the ranges of different LOS categories. Speed data used in this study are collected using Trimble GeoXT GPS receivers fitted on mid-sized vehicles for five major urban corridors comprising of 100 street segments of Greater Mumbai region. Result shows that FFS of urban street classes and average travel speed of LOS categories are lower than that evoked in HCM 2000 on account of physical and surrounding environmental characteristics. Also, average travel speed of LOS categories expressed in terms percentage of FFS of urban street classes found to be different from that mentioned in HCM 2010.

KEY WORDS

Urban Roads, Level of service (LOS), Clustering Analysis, adaboost algorithm, Genetic programming (GP), Maximum-Likelihood Method (ML), Expectation-Maximization Method, (EM) and Free Flow Speed (FFS).

TABLE OF CONTENTS

LIST OF FIGURES

Figure No. **Page No.**

LIST OF TABLES

Table No. **Page No.**

Chapter 1

Introduction

1.1 General

The simplest definition of urbanism or an urban area might be: confederation or union of neighbouring clans resorting to a canter used as a common meeting place for worship, protection, and the like; hence, the political or sovereign body formed by such a community. An urban area can also be defined as a composite of cells, neighbourhoods, or communities where people work together for the common good. The types of urban areas can vary as greatly as the varieties of activities performed there: the means of production and the kinds of goods, trades, transportation, the delivery of goods, and services, or a combination of all these activities. A third definition says that urban areas are those locations where there is opportunity for a diversified living environment and diverse lifestyle. People live, work and enjoy themselves in social and cultural relationships provided by these proximities of an urban area. Urban areas can be simplex or complex. They can have a rural flavour or that of an industrial workshop. They can be peaceful or filled with all types of conflict. They can be small and easy to maintain, or gargantuan and filled with strife and economic problems. According to a study, made by the Government of India, there were 83 cities in India by the end of 2003, with a population of over 0.25 million. Travel demand in these cities is 443-billion passenger kilometres. Nearly 80 percent of the demand was expected to be met by road based transport systems (Datla, 2004). At present no proper methodology is available to evaluate Level of Service (LOS) provided by urban streets in India. It is important to develop suitable methodologies for level of service analysis of urban streets. The Greek philosopher

Heraclitus once said that the difficulty confronting human society was to combine that degree of liberty without which law is tyranny, with that degree of law without which liberty becomes licence. The democracy of Athens, the Magna Carta in England, and the constitution of United States were wrought essentially from that same precept. Our society is organised on the basis of the group of laws established to guide the people in their conduct. The people of America created a vast domain of commercial and industrial enterprise and they built great cities in proper planning and defining the level of service criteria of roads. It is alarming to note that 32 percent of these vehicles are plying in metropolitan cities alone, which constitute about 11 percent of the total population (MORTH, 2003).

After independence, India's growth was based to assist the urban area by introducing the industrialization and urban infrastructure. Since urban agglomeration in India is the set of large urban clusters where the built up zones of influence the distinct cities or towns are connected by a continuous built-up development. But the Level of Service (LOS) analysis of urban roads in India is not decently delimited for majorly heterogeneous traffic flow. Subsequently India is a populated country, therefore heterogeneous traffic flow occurs in urban roads. The Level of Service (LOS) delineate for urban roads in India by HCM 2000, which is worthiest for homogeneous traffic flow. In reality, homogeneous traffic flow is outlined as a region of high density and low average velocity of cars and the flow of heterogeneous traffic on urban roads are extremely composite and the existing analytical model cannot be employed to auspicate the Level of Service (LOS) of urban roads. Thus an endeavour has been made to specify the Level of Service (LOS) criteria for India in this study.

The speed data were accumulated by Trimble GeoXT receiver, where the GPS receiver mounted on a vehicle and automatically records location of urban corridors and speed at regular sampling interval. The development of information technology and advancement of global positioning system (GPS) has largely overcome the data quality and quantity shortcomings of the manual and distance measuring instrument (DMI) methods of accumulating travel time data become one of the alternatives to the moving car observer method for the field data collection. . DMI measures the speed distance using pulses from a sensor attached to the test vehicle's transmission .This method also has some limitations like very complicated wiring is required to install a DMI unit to a vehicle. Frequent calibration and verification factors unrelated to the unit are necessary to store making the data file large and which leads to a data storage problem. The automated procedure provides convenience, consistency, finer precision and accuracy than the conventional procedure.

Defining LOS is essentially classification problem. The literature review suggests that cluster analysis is the well nigh desirable technique for the classification of the large amount of speed data germinated through GPS receiver. Four advanced clustering technique such as Adaboost algorithm, Genetic Programming, Maximum-likelihood algorithm and Expectation-Maximization method are employed for clustering intention in this study. Before going to the clustering, the optimal no of clusters were found out by applying five validation parameters on FFS data. Subsequently getting optimal no of clusters, these clustering algorithms were utilized twice in this research. First, four clustering algorithm were employed on Free Flow Speed (FFS) data to acquire speed ranges of urban street classes. Later delineate the speed ranges; again four clustering algorithms were enforced for the second time, on the average travel speed data to acquire the speed ranges of different LOS categories. The coherence of the clustering result of compartmentalization of urban streets and Level of Service (LOS) categories were assured with geometric and surrounding environment features of street segments. The overall model of this study is as shown in the figure 1.1.

The overall model of this study is illustrated in Figure 1.1

Selection and Delineation of the Study Area i.e Road Networks & Study Corridors

⇓

Geo-XT GPS Receiver fitted with Mid-Sized Vehicles for Data Collection

⇓

Data Accumulation

- Free Flow Speed
- Average Travel Speed during both Peak and Off-Peak hours
- Physical and Surrounding Environmental Features

⇓

GPS Tracks and Signal Cleansing

- Cleansing of Scatter Type Data
- Join with Survey Data
- Interpolation
- Final Data Set

⇓

Interpolation and Comparison

- FFS & MOE in terms of Average Travel Speed
- Cluster Validation
- Suitable Clustering Algorithm
- Urban Street Classification
- Speed Ranges Of LOS Categories
- Clustering Result Comparison with Physical and Surrounding Environmental Characteristics

⇓

Summary, Conclusion, Limitation and Future work

Figure-1.1: Overall model of the study

1.2 Statement of the problem

Urban Street is a complex organism. It is a great human enterprise that should serve the material and spiritual needs of humanity. The city is mosaic of homes and shops, factories and offices, schools and libraries, theatres and hospitals, parks and religious institutions and also the urban street. The urban street is seriously suffering for decreasing speeds, increased congestion, increased travel time, and decreased level of service and increase in accident rates.

Floating car method was used traditionally for collection of travel speed data. Although this method is very simple but it has some lacuna like accuracy variation from technician to technician and the possibility of missing and inaccurate marking of some check points. Recent research has demonstrated the feasibility of using Global Positioning System (GPS) and Geographic Information System (GIS) technologies for automating the travel time data collection, reduction, and reporting when using a probe vehicle. The new automated procedures provide consistency, automation, finer levels of resolution, and better accuracy in measuring travel time, delay and speed rather than traditional techniques. As a result, large amounts of reliable travel time, delay and speed data can be collected and processed.

The LOS for urban streets is very important for analysis of urban streets. The LOS affects the planning, design, and operational aspects of transportation projects as well as the allocation of limited financial resources among competing transportation projects. Traffic composed of identical vehicles and following lane discipline is termed as homogeneous traffic. Traffic with the presence of motorized and non-motorized two-wheelers and three-wheelers along with several other vehicles with no-lane discipline, is termed as heterogeneous. This heterogeneous traffic is clearly different from the one with the presence of trucks and has also been termed as heterogeneous traffic. Since India is a developing country and no suitable method yet to describe for heterogeneous traffic in urban Indian context.

1.3 Objectives and scope

Based on the above problem statement, the objectives of the study are:

- To classify the urban street segments into various classes using free flow speed data acquired through GPS and data clustering technique.
- To define free flow speed ranges of urban street classes and speed ranges of LOS categories using advanced clustering algorithms.
- To find the most suitable cluster analysis algorithm in defining FFS ranges of urban street classes and speed ranges of LOS categories.

1.4 Organization of report

The report describes about six chapters. The first chapter compromises introduction to research work and detail description about statement of the problem, objectives and scope of the study. The second chapter is literature review part which includes the level of service, use of GIS-GPS in traffic data collection and clustering techniques. Third chapter provides idea about the study area of this work and methodology of data collection. The forth chapter is for detail description on cluster techniques. Result and analysis for the research work is found out in the chapter five. The summary and conclusion are in the chapter six and it also includes the limitation and scope for the future work.

Chapter 2

Literature Review

2.1 General

The level of service (LOS) concept was first introduced in 1965 version of highway capacity manual. In this concept, it was recognized that the driver's view of the transportation system is also important to consider. It has become more important to estimate not only the LOS but also key operational performance measures like queue length or average speed. it has become important to expand the analysis area from a single point to a segment, and then form a linear segment to a two dimensional area. And then ultimately on found a single, integrated multimodal transportation system. The 1950 HCM was prepared only to fulfil the needs of traffic engineers who were participating in planning, design, and handling specific roadway components. According to 1965 HCM, level of service described by six classes from "A" to "F" defined based on the combination of travel time and the ratio of traffic flow rate to the capacity road sections. The "1965 HCM' concept was redefined to several traffic conditions in the 1985 version of highway capacity manual include travel speed, traffic flow rate and traffic density of each type of roads. There was limitation to 1985 HCM that LOS measure which was given by Baumgartner (1996), Cameron (1996) and Brilon (2000). Baumgartner (1996) realized the rapid growth of urban populations, ownership of vehicle, trip length, and number of trips has resulted in a relative increase in traffic volumes. Thus, travel condition that would have been viewed as intolerable in the 1960s are considered normal by today's motorists, especially commuters Cameron (1996) stated that it was not uncommon to wait three minutes as a congested urban intersection with average delays often exceeding two minutes. Later various researches have been done for describing six level of service designations to nine or more.

The LOS criterion of roads in urban street is basically classification problems and cluster analysis is a suitable technique for that classification. Large amount of free flow speed (FFS) and average travel speed data are needed for cluster analysis because LOS of urban street is a function of travel speed along street segments. The travel speed data were collected by using floating car method. Turner et. al. (1998) found that the method has produced the susceptible human error. Later distance measuring instrument (DMI) was introduced as a solution of floating car method. Benz and Ogden (1996) found limitation to this method because it has produced difficulties in DMI unit and data storage problems. The HCM describes "level of service" is a qualitative measure that describes traffic conditions in terms of speed, travel time, freedom to manoeuvre, comfort and convenience, traffic interruptions and safety. Six classifications are used to define LOS, designated by the letters "A" to "F". Where LOS A represents the best conditions, while LOS F represents heavily congested flow with traffic demand exceeding highway capacity.

The current definition of LOS being followed is that defined in HCM 2010 "LOS is a quantitative stratification of a performance measure. The measure employed to determine LOS for transportation system elements are called "service measures". For heterogeneous traffic condition in India, Marwah and Singh (2000) have classified LOS into four groups (I-IV). Similarly, Maitra et al. (1999) taking congestion as measure of effectiveness for prevailing heterogeneous traffic condition in India divided LOS into nine groups "A" to "I". Baumgaertner (1996) pointed out that today motorist became more adapted to urban congestion so the traffic condition which was viewed as intolerable 1960 now considered normal. Kita and Fujiwara (1995) stated LOS not to be a traffic operating condition but tried to find the relationship of LOS with driver's perception. Spring (1999) found service quality to be a continuous and subjective matter so the author opined it to be inappropriate to have distinct boundary or threshold value for a particular LOS. Flannery et al. (2005) suggested LOS does not completely represent drivers' assessment on performance of urban streets. The author opined for inclusion of qualitative measures for defining LOS. Clark (2008) from a study upon the prevailing traffic condition of New Zealand suggested for a new LOS category to be termed as F+ or G.

Shao and Sun (2010) categorized LOS into two parts: Level of facility supply and Level of traffic operation. Travel speed to free flow speed ratio was considered as evaluation index of traffic operation. Traffic operation categorized into different groups using Fuzzy set. Determination of LOS of Urban street from user perception was carried out by Flannery et.al (2008). Fang et. al. (2011) determined speed-flow curves of different segments of an interchange by developing a simulation model using VISSIM software. LOS ranges from the speed-flow curves were determined by taking density as classification index. Arasan and Vedagiri (2010) through computer simulation studied the effect of a dedicated bus lane on the LOS of heterogeneous traffic condition prevailing in India. A state-of-the-art hybrid algorithm was developed by Ivana et.al. (2011) to classify urban roads based on vehicle track and infrastructural data collected through GPS. From the study the authors found the limitations of traditional clustering algorithm in classifying large amount of speed data. Not going with traditional research in which traffic flow is considered as the only parameter to access the LOS of traffic facilities, Tan et. al. (2007) analyzed the pedestrian LOS with physical facilities and traffic flow operation along with user perception. Limitations of LOS criteria of walkways proposed by HCM 2000 for China were found out by Cao et.al. (2009). User perception taken into consideration for classification of LOS at urban rail transit passages and found the limit for LOS standards suitable for China is lower than that suggested by HCM 2000. Body size, culture, gender and age of user found to be influence factors for the LOS classification.

Bhuyan and Rao (2011) defined the free flow speed ranges of urban street classes and speed limits of LOS categories using Hierarchical Agglomerative Clustering (HAC) and data collected by GPS handheld receiver in Indian context. For developing countries, Maitra et.al. (2004) have shown the effect of different types of vehicles on congestion trough congestion model. The established model can be used as a tool for formulating traffic management measures for urban roads. Basuet.al. (2006) modelled passenger car equivalency for urban mid-block using stream speed as measure of equivalence. In this study a neural network approach was explored to capture the effects of traffic volume and its composition level on the stream speed. Chung (2003) tried to determine the travel pattern along a particular route of Tokyo metropolitan area. Kikuchi and Chakroborty (2007) utilized Fuzzy set in order to find the uncertainty associated with the LOS categories. Six frameworks were proposed by the authors in order to determine the uncertainty associated under each LOS category.

Shouhua et.al. (2009) found the LOS criteria of walkways proposed by HCM 2000 are not suitable for China. The authors have taken user perception into consideration for classification of LOS at urban rail transit passages and found the limit for LOS standards suitable for China is lower than that suggested by HCM 2000. Fang and Pechuex (2009) studied about the LOS of a signalized intersection taking user perception into account. The author found that it is best to differentiate LOS into six categories as described in HCM but proposed a new six LOS by merging existing LOS A and B and splitting existing LOS F into two categories. Pattnaik and Ramesh Kumar (1996) developed methodology to define level of service of urban roads taking into account users' perceptions. Kittelson and roess (2001) have noted down that the HCM (2000) methodologies have not been based upon user perception surveys. The HCM(2000) methodologies have resulted from a combination of consulting studies, research, debates and discussions of the highway capacity and quality of service(HCQS) committee(Pecheux et.al,2000). In July 2001, at the midyear meeting of the HCQS committee, a motion was passed that stated “the committee recognizes that there are significant issues with the current LOS structure and encourages investigations to address these issues”. Brillion and Estel (2010) have presented standardized methods that allow a differentiated evaluation of saturation of flow (LOS F) conditions beyond static considerations of traffic conditions in German highway capacity manual. According to Indian roads congress(IRC 1990) , for an urban roads, the LOS are strongly affected by factors such as heterogeneity of traffic, speed regulations, frequency of intersections, presence of bus stops, on-street parking, road side commercial activities and pedestrian volumes etc.

The level of service (LOS) concept differs from country to country. In North America the LOS is of six types. i.e. from “A” to “F”.

A= free flow

B= reasonable free flow

C= stable flow

D= approaching on stable flow

E= unstable flow

F= forced or break-down flows.

The LOS concept also suited in the country UK and Australia. In Australia the LOS are an integral component of asset management plans. The HCM 2010 version corporate tools for multimodal analysis of urban streets. The primary basis for the new multimodal LOS procedures on urban streets is NCHRP report 616: multimodal level of service of analysis for urban streets. The researches have been developed for evaluating multimodal level of service (MMLOS) provided by different urban street designs and operations. The researchers can use the (MMLOS) to evaluate various street designs in terms of their effects on the auto driver's, transit passengers, bicyclist, and pedestrian's perceptions.

2.2 Methods of Cluster Analysis

Cluster analysis divides the data into conceptually meaningful groups of objects that share common characteristics. The meaningful groups are the goal, and then the clusters should capture the natural structure of the data. Clusters are the potential classes rather than it also useful for data summarization and automatically finding classes. Various clustering methods have been applied in speed data for creating the meaningful groups. There are four types of clustering algorithm have been applied for this research work. Such as: - Adaboost algorithm, Genetic programming, Maximum likelihood method and Expectation-Maximization Method. These four algorithms are useful for classification of urban streets and LOS categories.

2.2.1 Adaboost Algorithm

Adaboost was introduced in 1995 by Freund and Schapire. It is a well known large margin learning algorithm that can select a small set of the most discriminative features and combines them into an ensemble classifier and also an additive model. Viola-Jones's work (robust real time object detection) made adaboost learning world focus in the community of computer vision and pattern recognition. It is a Meta algorithm and can be used in conjunction with many other learning algorithms to improve their performance. Adaboost is a adaptive in the sense that subsequent classifier built area tweaked in favour of those instances learning algorithm.

Several adaboost variants have been proposed in the literature such as discrete boost (viola and Jones 2004), Bayesian boost (Xiao et.al, 2007), Real boost (Fried-manetal, 2000), KL boost(liu and shom,2003) try to minimise the some kind of classification error . so, that the selected feature is discriminative .however their stump based component classifier loss much discriminative information. In contrast, confidence-rated boosting (Fried manetal, 2000), (Lieu and Shum, 2003), (Schapire and Singer 1999) evolves with LUT –based component classifier. However their weak learners (e.g.: KL divergence) such as (Liu and Shum, 2003) cannot necessarily minimise the classification error. This means un necessary loss of efficiency of the weak learner in the view point of classification error.

The AdaBoost algorithm (2012) iteratively works on the Naïve-Bayesian classifier with normalized weights and it classifies the given input into different classes with some attributes. AdaBoost algorithm (2010) immediately minimizes the classification error of each selected feature, and thus activates the final detector to be more discriminative and to converge more quickly. A new method for license plate detection based on AdaBoost is proposed by Hauchun tan et. al. Adaboost algorithm (2009) using clustering and boosting to prune Bagging ensembles is proposed and Its learning efficiency is close to Bagging and its performance is close to AdaBoost., this algorithm can detect noisy data from original samples based on cascade technique and a better result of noise detection can be obtained.
The algorithm reflects the strong back ground for valuable research in the area of level of service criteria of roads in urban Indian context.

2.2.2 Genetic Programming

Genetic Programming is a collection of methods for the automatic generation of computer programmes that solve carefully specified problems. it is the initially random computer programmes, where only the relatively more successful individuals pass on the genetic material to the next generation.

Genetic programming was first used by Niles aall barricelli in 1954 which is applied to evolutionary simulations. the evolutionary simulation was registered as optimization methods in between 1960-1970. Ingo rechenberg and his group were solved various complex engineering problems by genetic programming. John holland was influence in 1970s. In 1964

Lawrence j.fogel discovered finite state automata of Genetic programming methodology in 1964. Later the learning classifier system developed by Genetic programming, which developed sets of sparse rules describing optimal policies for Markov decision process. The "tree based" genetic programming was developed by Cramer (1985) and it also extended by main proponent of Genetic programming .koza. Later various researches have been done for solving complex optimization. Gianna giavelli developed the Genetic programming as a technique to model DNA expression. Then, Genetic programming was mainly used to solve various simple problems. It has been applied to evolvable hardware as well as computer programme. The (schema theories, Markov chain models and meta-optimization algorithms) has been developed by Genetic programming. Genetic programming is a systematically domain independent method for getting computers to solve problems automatically starting from a high level statement. Pedro et.al. (2010) surveys existing literature about the application of Genetic programming (GP) to classification to show the different ways in which this evolutionary algorithm can help in the construction of accurate and reliable classifiers. Shamir et.el. (2006) purports an algorithm called parallel hybrid clustering using Genetic programming and multi-objective fitness with density (PYRAMID). PYRAMID employs a combination of data parallelism in the form of genetic programming and a multi-objective density based fitness function in the context of clustering to resolve most of the challenges such as: - identifying clusters of arbitrary shape, sensitivity to the order of input and dynamic determination of the no. of clusters.

2.2.3 Maximum Likelihood Method

The parameters of a statistical model is estimates by Maximum-Likelihood Method. When applied to given data set of a statistical model, maximum-likelihood estimation provides estimates for the model's parameter.

Maximum-likelihood estimation was first developed for Bayesian statistics, and then simplified by later authors. Maximum-likelihood estimation was recommended, analyzed (with flawed attempts at proofs) and vastly popularized by R. A. Fisher between 1912 and 1922. Although it had been used earlier by Gauss, Laplace, T. N. Thiele, and F. Y. Edge worth. In 1922 R. A. Fisher introduced the method of maximum likelihood & first presented the numerical procedure in 1912. There are two general methods of parameter estimation. They are least-squares estimation (LSE) and maximum likelihood estimation (MLE). Maximum likelihood estimation is requirement for the chi-square test, the G Square test, Bayesian methods, illation with missing of data, modeling of random effects, and many model selection criteria such as the Akaike information criterion (Akaike, 1973) and the Bayesian information criteria (Schwarz, 1978).

Xiong (2009) analyses the statistical relationship between average headway and capacity. The average headway distribution based on the MLE. Then, find the changes of average headway with occupancy locate the minimum of average headway and thus obtained the road capacity. The critical gaps and follow-up times were analyzed by observing traffic flow at German territory intersection by weinert to determine their dependence on parameters such as intersection layout, speeds and volumes by applying maximum likelihood method. Hang (2009) applied the maximum likelihood clustering algorithm to separate road from other ground objects. The road traffic and lane lines are advance detected by using texture enhancement and morphological operations after the detection of road surface. The maximum likelihood method proposed by Ernst. et.al. (2009) for significantly improved speed estimates that can used to produce histogram of vehicle speeds instead of the speed averages. The fuzzy classifier makes use of spatial features extracted from a multi spectral data and a classification image is generated by maximum likelihood method which has studied by shivakumar et.al. (2013).

2.2.4 Expectation-Maximization Algorithm

In statistics, an Expectation–Maximization (EM) algorithm is an iterative method for finding maximum likelihood or maximum a posterior (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the *E* step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

The EM algorithm was explained and given its name in a classic 1977 paper by Arthur Dempster, Nan Laird, and Donald Rubin. In particular the EM method for exponential families was published by Rolf Sundberg in his thesis and several papers following his collaboration with and Anders Martin-lof .The “Expectation-Maximization" (EM) algorithm is a general technique for maximum likelihood (ML) or maximum a posterior (MAP) estimation. The recent emphasis in the artificial neural network literature on probabilistic models has resulted to increased interest in EM as a possible alternative to gradient-based methods for optimization. The empirical results reported in these papers suggest that EM has considerable promise as an optimization method for such architectures.

Anditsai et.al. (2005) proposed an EM algorithm for estimating the shape contours that illustrate the different shape classes. Kayabol and koray (2011) proposed unsupervised classification of SAR (synthetic aperture radar) images using EM algorithm. Un supervised clustering of normal vessel traffic patterns is proposed and implemented by laxhammar. et. al. (2008) where patterns are represented as the momentary location, speed and course of tracked vessels. Expectation-Maximization algorithm is used as a clustering algorithm in which Gaussian mixture model are used for anomaly detection in sea traffic. Chen et. al. (2009)

analyzing the traffic flow data in china to evaluate the highway region's traffic operation state so as to improve traffic control strategy in which EM algorithm used to extracting an eigen vector to describe highway traffic flow state of similar region section.

2.3 SUMMARY

From the literature review, it was found that; the LOS is not passable defined for heterogeneous traffic flow with different operational features in Indian context. GPS was found to be accurate technique for collecting speed data. Delineating LOS is essentially classification problem. Clustering analysis is the worthiest proficiency to solve the classification problem. The next chapter gives the detail idea about study area and data collection technique.

Chapter 3

Study Area and Data Collection

This section dissevered into two parts. The first part concisely depicts the study corridors from where the speed data as well as the road inventory data were accumulated. The second part clarifies the contingents of data accumulation technique espoused for this study.

3.1 Study Corridors

Five important urban road corridors of the commercial city Mumbai of Maharashtra state, India are picked out for this study. Greater Mumbai is an island city with a linear pattern of transport network having prevail north-south commuter movements. Passengers move towards south for work trip in the morning hours and return back towards north in the evening hours. Thence four north-south corridors and one east-west corridor comprising of 100 street segments have been preferred for this study. Major roads like eastern express highway broadening up to south(corridor-1), LBS road diversifying up to south via Ambedkar road (corridor-2), western express highway stretching up to marine drive(corridor-3), SV road covering up to south via veer Savarkar road (corridor-4) and Versova-Andheri-Ghatkopar-Vashi (VAGV)(corridor-5) are considered. These are shown on the GIS base map in Figure 3.1.

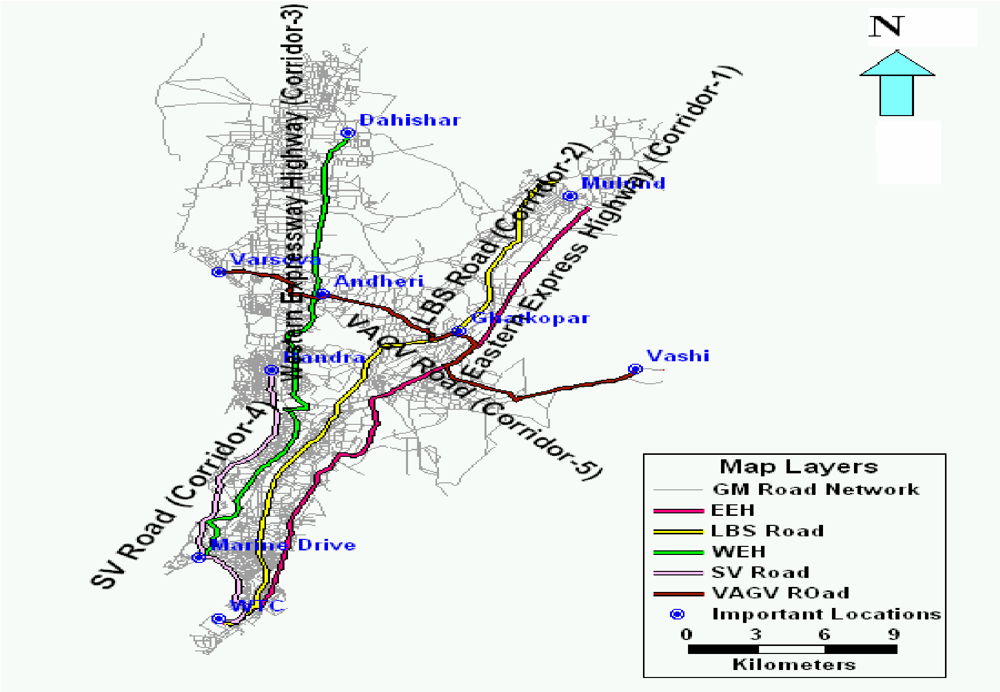


Figure- 3.1: Map showing selected corridors of Greater Mumbai

3.2 Data Collection

The probe vehicles featuring mid-sized car was tallied with Trimble Geo-XT GPS receiver, adequate to logging speed data unremittingly at time intervals of one second. GPS furnishes both spatial and time/distance based data from which various traffic parameters were deduced, letting in travel time and travel speeds. In order to bring forth unbiased data sets three mid-sized cars were employed and assist of three drivers on different days of the survey work was chosen. Essentially three types of data sets were accumulated.

The first type is roadway inventory contingents, for which a data dictionary was geared up using path finder office 3.0. Throughout the accumulation of inventory contingents, proper segmentation technique was implemented, which is just afterwards signalized intersection to just afterwards next signalized intersection. Contingents on segments like segment number, number of lanes, median types, pedestrian activity, road side development, access density, construction activity, speed limit, separate right turn lane, number of fly overs, date and day of data accumulation and segment length were accumulated.

The second type of survey conducted was to break through the free flow speed. Before going for the free flow speed data collection, we should know when the traffic volumes is less than or equal to 200 vehicles per lane per hour. An elaborated 24-hour traffic volume count survey was carried out by this group for western sea-link (WFSL) project. The traffic volume data were accumulated on 45 stations on seven screen lines. From these survey data traffic volumes per lane per hour was deliberate for roads comings under this study area. It was established that free flow traffic condition (less than 200 veh/In/hr) is forthcoming at 12 mid-night and all road sections are having free flow traffic conditions from 1AM to 5AM. Thus free flow speed for all these corridors were accumulated during these hours.

The third type of data accumulated was congested travel speed. Congested travel speed survey was comported throughout both peak and off pick hours on both directions of all corridors. Number of trips extended for each direction of travel and for the study hours (peak, off-peak and free flow) is at least 3 and sometimes it is up to six trips. Afterward data has been accumulated in the field; it has been transported back to the office computer by using path finder office version 3.0. Accuracy of field data were significantly ameliorated through a process called differential correction.

3.3 Summary

This chapter provided the details of the study area, data collection and database preparation. The details of corridors on which GPS data was collected were discussed. Pathfinder was used to prepare data dictionary and the inventory details were collected using the prepared data dictionary. It was also discussed how the timing of free-flow speed data collection was fixed based upon the traffic volume data.

The next chapter gives idea about the cluster analysis algorithms used in this study and also about the various cluster validation parameters used in the research work in order to determine the optimal number of cluster and to select the best clustering algorithm.

Chapter 4

Cluster analysis

4.1 Cluster Analysis

Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval and bioinformatics. Cluster analysis groups data objects based on only information found in the data that describes the objects and their relationships. The goal is that the objects within the group are similar to one another and different from objects in the group. Different types of cluster analysis are there. In this research, there are four methods are used. Popularly, known as Adaboost algorithm, Genetic programming, Maximum likelihood Method and Expectation-Maximization algorithm.

4.2 Adaboost Algorithm

Adaboost algorithm was introduced in 1995 by Freund and Schipre. It is a well known large margin learning algorithm that can select a small set of most discriminative features and combines them into an ensemble classifier. Viola-Jones's work made Adaboost learning world focus in the community of computer vision and pattern recognition.

The mechanics of adaboost learning algorithm includes three fundamental points such as weak learner, the component classifier and the re-weighting function.

- **Weak learner:** The weak learner is essentially the criterion for choosing the best feature *-t* on the weighted training set.
- **The component classifier:** The component classifier outputs the confidence of a sample being a positive based on its t value.
- **Sample re-weighting:** Sample re-weighting enables that the subsequent component classifier can concentrate on the hard samples by assigning higher weights to the samples that are wrongly classified by previous component classifiers.

Adaboost is one of the most influential ensemble methods. Its birth was originated from the answer to an interesting question posed by Kearns and Valiant (1988). That is, whether two complexity classes, weakly learnable and strongly learnable problems, are equal. Adaboost and its variants have been applied to diverse domains with great success, owing to their solid theoretical foundation, accurate prediction and great simplicity. The adaboost algorithm was described by Viola and Jones (2001).

- Step 1: Given a set of training samples (x1,y1),(x2,y2)…..(xn,yn) where yi = 0 for negative sample, yi = 1 for positive sample. N is number of total training example.
- Step 2: Initialize weights W1,i = D(i), for negative D(i) = 1/(2m), where m is number of negative samples. For positive D(i) = 1/(2l), where l is number of positive samples. m + l = N.
- Step 3 : for t= 1 to T,

  A= normalise the weights

$$q_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}} \tag{4.1}$$

  B= For each feature, f, train a classifier h(x, f, p, θ) The error is evaluated with respect to qt:

$$\varepsilon_f = \sum_i q_i \left| h(x_i, f, p, \theta) - y_i \right| \tag{4.2}$$

C: Choose the classifier, ht, with the lowest error **ε:**

$$\varepsilon_t = \min_{f,p,\theta} \sum_i q_i \left| h(x_i, f, p, \theta) - y_i \right| = \sum_i q_i \left| h(x_i, f_t, p_t, \theta_t) - y_i \right| \tag{4.3}$$

$$h_t(x) = h(x, f_t, p_t, \theta_t) \tag{4.4}$$

D: Update the weights:

$$h(x) = \{1 \sum_{t=1}^{T} \alpha_t h_t(x) \geq 1/2 \sum_{t=1}^{T} \alpha_t \ \textit{and}\ 0 \text{ otherwise.} \tag{4.5}$$

Where $\alpha_t = \log \frac{1}{\beta_t}$ (4.6)

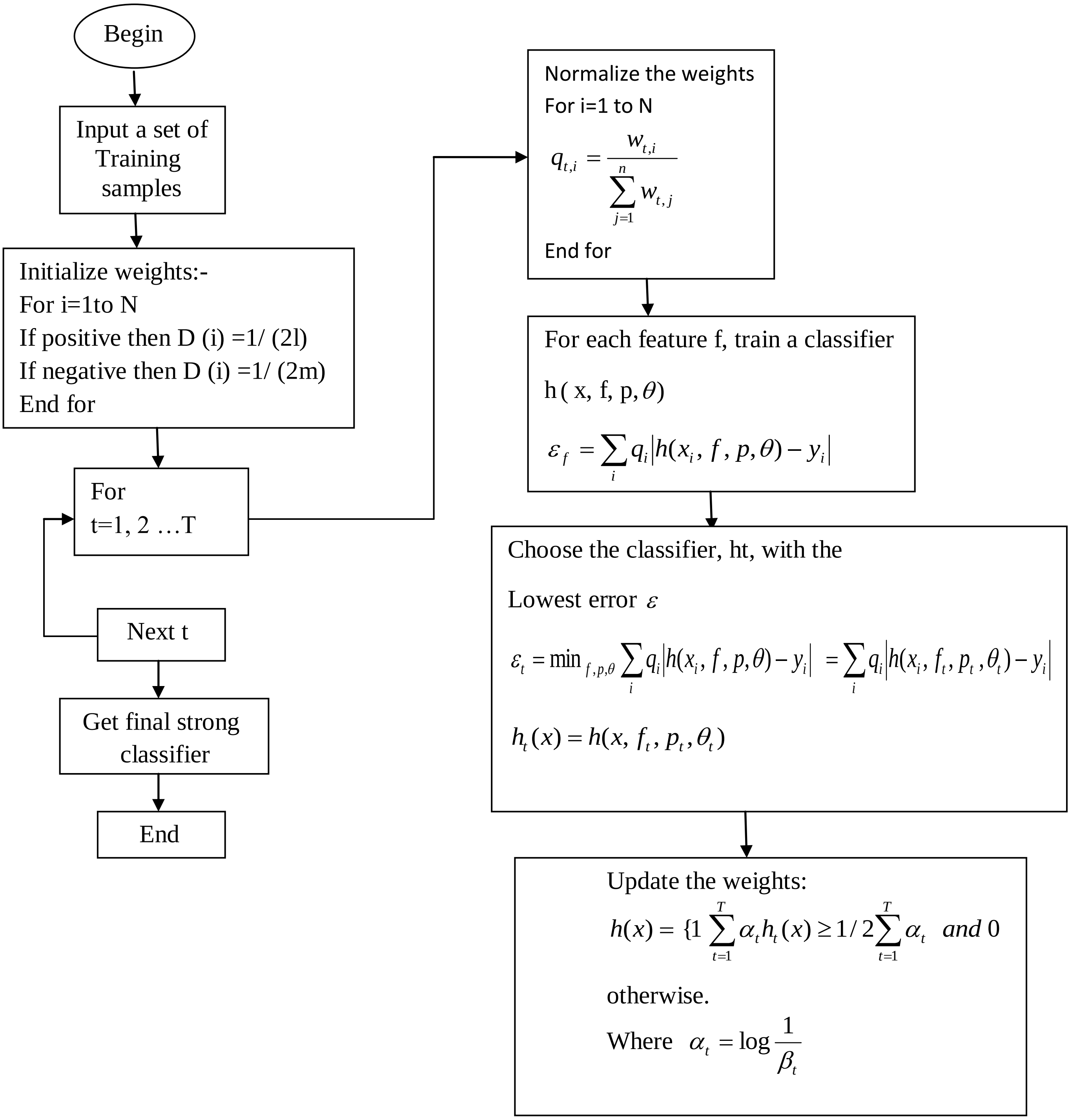


Figure-4.1: Flow chart of adaboost algorithm

4.3 Genetic Programming

Genetic programming is a powerful method for automatically generating computer programs via the process of natural selection. It uses a genetic algorithm to search through a space of possible computer programs for one which is nearly optimal in its ability to perform a particular task. Genetic programming's unique representation of solution distinguishes it from genetic algorithm. In GA, solutions are typically represented as bit strings, or genotypes, with predefined regions of the bit string mapped to specific phenontypic traits. In Genetic Programming, solutions are represented directly as computer programmes, written in some suitable domain specific language. The flow chart of the methodology followed for genetic programming in this study is shown in Figure 4.2.

4.3.1 Genetic Programming Algorithm

The Genetic Programming algorithm is a domain independent method. It provides a single unified approach to the problem of finding a computer program to solve a problem. In summary, the Genetic Programming algorithm breeds computer programs to solve problems by executing the following five steps.

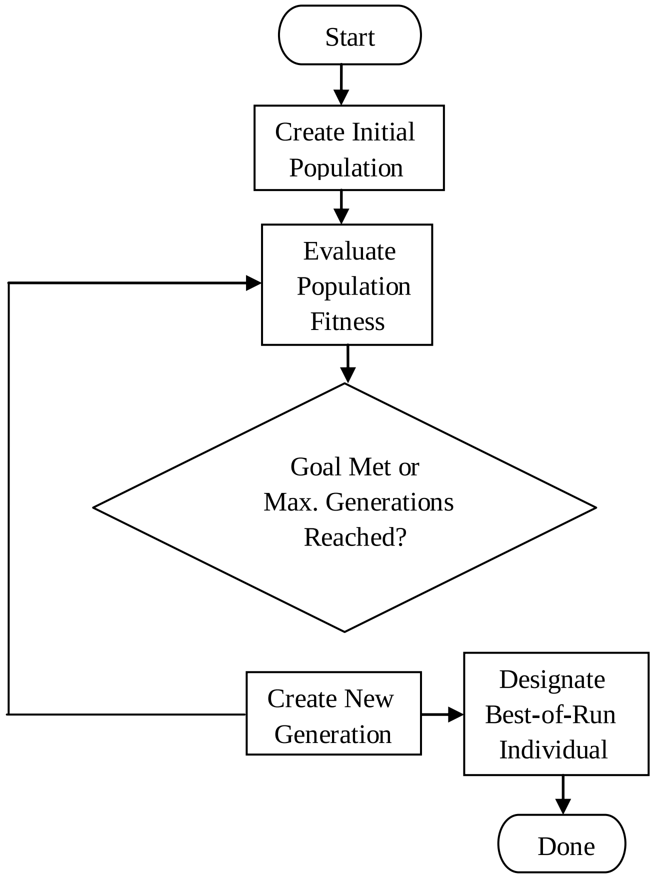


Figure-4.2: Genetic programming flow chart

Step-1: -Generate **(**Gen=0) an initial population of random composites of the function and the terminals of the problem.

Step-2: -Then candidate's fitness is measured by the fitness function. Once candidates are evaluated, each one's raw fitness is converted to a normalized fitness via three step process.

a. **Calculate standard fitness: -** standard fitness is simply the raw fitness expressed such that all fitness values are positive and lower fitness values correspond to better performance. The best possible fitness if known, then fitness = 0.

b. **Calculate adjusted fitness:-** adjusted fitness is defined as a

$$a(i) = \frac{1}{1+s(i)} \tag{4.7}$$

Where $s(i)$ is the standardized fitness for individual i in this case, larger values denote better performance.

c. **Calculate normalized fitness**:-Normalized fitness is defined as

$$n(i) = \frac{a(i)}{\sum_{K=1}^{M} a(k)} \tag{4.8}$$

Where $a(i)$ is the adjusted fitness of individual i and M is the population size. In this case larger values denote better performance.

Step-3:-

a. Find a function that best satisfies a set of fitness cases generated by the quadratic polynomial function.

$$f_1(x) = x^4 + x^3 + x^2 + x \tag{4.9}$$

b. Find a function that best satisfies a set of fitness cases generated by the function.

$$f_2(x) = \sin(x^4 + x^2) \tag{4.10}$$

c. Find a function that best satisfied a set of fitness cases generated by the function.

$$f_3(x) = \sin(\exp(\sin(\exp(\sin(x))))) \tag{4.11}$$

Step-4:-Let I be the only genetic operator is recombination and no mutation is acting. Let β be the space of different blocks of code. Let $C(i/j)$ be the probability that the recombination operator picks out a particular block i from a program j . So

$$\sum_{i \in \beta} C(i/j) = 1, \text{ For all } j \in S \tag{4.12}$$

Then the frequency p_i, that the operator picks block i from a random program in the population is

$$p_i = \sum_{j \in s} C(i/j) x_j, \; i \in \beta \tag{4.13}$$

Step-5:-The best computer program that appeared in any population (i.e the best-of-run individual) is designated as the result of Genetic Programming. The set of possible structures in genetic programming is the set of N_{func} functions from $F = \{f_1, f_2, ..., f_{Nfun}\}$ and the set of N_{term} terminals from $T = \{a_1, a_2, ..., a_{Nterm}\}$.

4.4 Maximum Likelihood Method

Maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.For some models the maximum likelihood parameters can be identified instantly from the data; for more complex models, finding the maximum likelihood parameters may require an iterative algorithm. For any model, it is usually easiest to work with the logarithm of the likelihood rather than the likelihood, since likelihoods, being products of the probabilities of many data points, tend to be very small. Likelihoods multiply; log likelihoods add. Estimation of parameters is a fundamental problem in data analysis. A handful of estimation methods existed before maximum likelihood, such as least squares, method of moments and Bayesian estimation.

4.4.1 Maximum Likelihood Method Algorithm

The maximum likelihood method describes the training data $\{X_j\}$ and cluster centers V_i.

Where $j = 1.......n$ and $i = 1........k$.

The other variables:

$h(i/X_j)$ =Probability for X_j to be in the ith cluster.

P_i =priori probability of selecting ith cluster.

F_i =covariance of the ith cluster.

V_i =cluster center of the ith cluster.

For running the algorithm the variables are initialize as:

$$h(i/X_j) = 1, i = \arg\min d(X_j, V_J) \quad i = 1.......k \tag{4.14}$$

$$P\text{i}= \sum_{J=1}^{N} h(\frac{i}{Xj})/N \tag{4.15}$$

$$F\text{i}=\sum_{j=1}^{N} h\left(\frac{i}{Xj}\right)(Xj - Vi)(Xj - Vi)^T/N \tag{4.16}$$

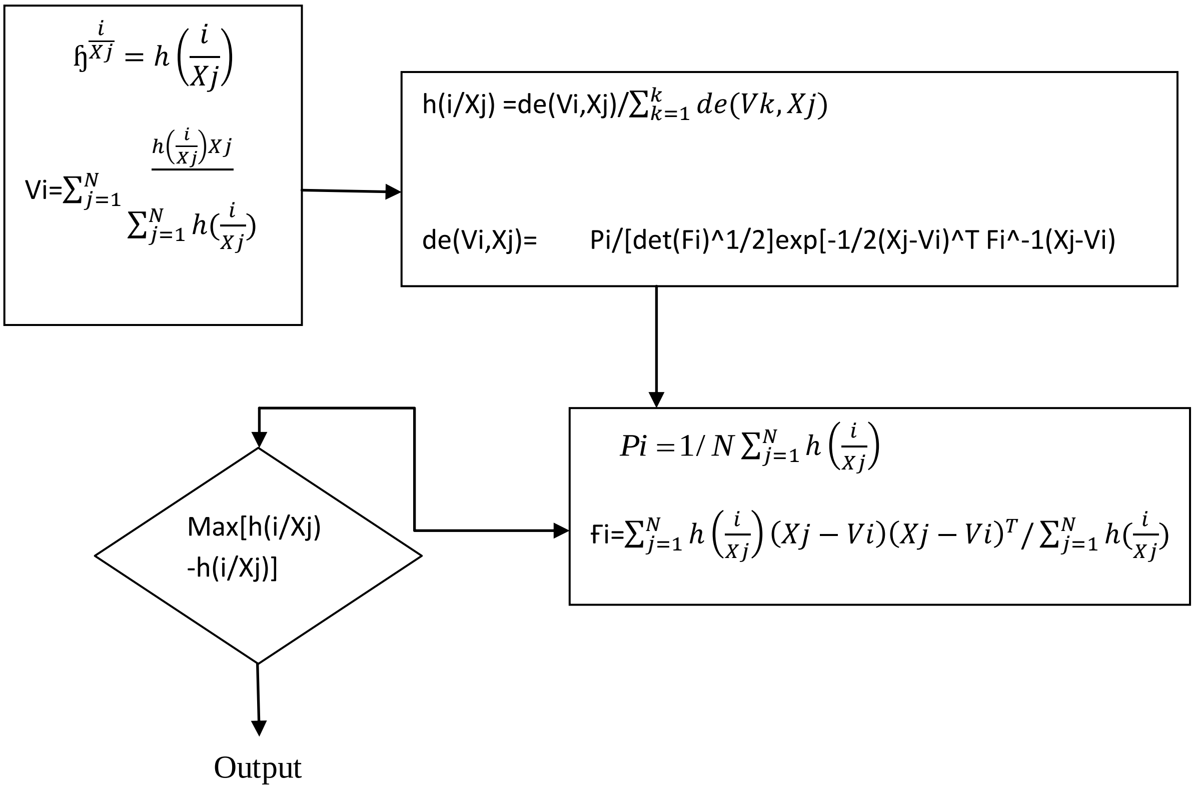


Fig 4.3: Maximum Likelihood Method algorithm

After the training phase, all the clusters representing the training data have been determined. now we are able to determine whether an object is similar to target or not. Then, first calculate the minimum distance between the object's feature vector and clusters. Then put a threshold on this distance. If it is a less than a predefined threshold, then it is called as a target or otherwise called as a clutter.

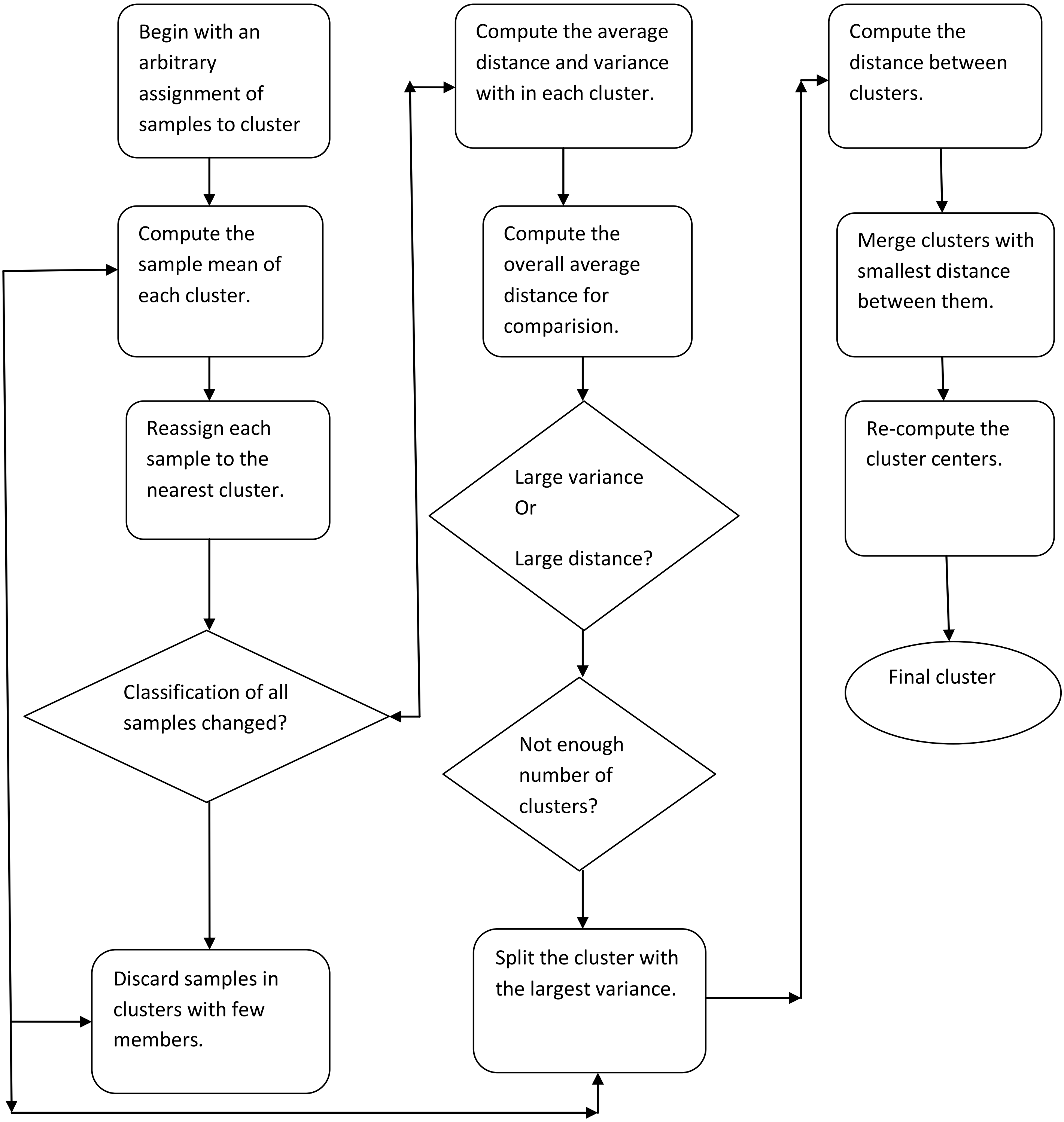


Fig 4.4: Flow chart of ML Method

4.5 Expectation-Maximization Algorithm

The EM algorithm is efficient iterative procedure to compute the maximum likelihood estimates in the presence of missing or hidden data. Each iteration of the EM algorithm consists of two processes i.e. E-step and M-step. For the expectation E-step, the missing data are estimated given the observed data and current estimate of the model parameters. In the M-step, for the maximization, the likelihood function is maximized under the assumption that the missing data are known.

Let $D \underline{\underline{\Delta}} \ (x_1, \ldots\ldots\ldots, x_N)$ be the observed data, and let $Z \underline{\underline{\Delta}}$ hidden random variables.

Let $\theta \underline{\underline{\Delta}}$ be the model parameters.

Then $\hat{\theta} = \arg\max_{\theta} \log p(x, z/\theta)$

$$= \arg\max_{\theta} \log p(z/\theta) + \log p(x/z, \theta) \tag{4.17}$$

The expression being maximized on the last line is known as the complete log likelihood.

In the latent setting:-

$$\hat{\theta} = \arg\max_{\theta} \sum_{z} p(x/\theta)\, p(x/z, \theta) \tag{4.18}$$

EM algorithm is usually described in two steps i.e. E-step and M-step. But here let it break down to five steps.

Step-1:- Given a training data set:-

$$\begin{aligned} X &= \{x(1), x(2), \ldots\ldots\ldots\ldots x(n)\} \\ Z &= \{z(1), z(2), \ldots\ldots\ldots\ldots\ldots z(n)\} \end{aligned} \tag{4.19}$$

$z(i)$ is the class label of sample $x(i)$.

Step-2:- Create a model by specifying a joint distribution

$$p(x(i), z(i)) = p(x(i)/z(i))\, p(z(i)), \tag{4.20}$$

Then parameters of the model ϕ, μ, Σ.

Step-3:- Begin with a guess for ϕ, μ, Σ and then iterate between expectation and maximization to improve the estimates of ϕ, μ, Σ, z **.**

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p(x^{(i)}/z^{i}; \mu, \Sigma) + \log p(z^{(i)}; \phi) \tag{4.21}$$

Maximizing this w.r.t. ϕ, μ, Σ gives the parameters

$$\phi_j = \frac{1}{m}\sum_{i=1}^{m} 1\{z^{(i)} = j\}, \tag{4.22}$$

$$\mu_j = \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{z^{(i)} = j\}}, \tag{4.23}$$

$$\Sigma_j = \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\}(x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} 1\{z^{(i)} = j\}}$$

(4.24)

Step-4:- Repeat until the convergence: (E-step)

For each i, j set $\omega_j^{(i)} = p(z^{(i)} = j / x^{(i)}; \phi, \mu, \Sigma)$ (4.25)

Step-5:- Update the parameter (M-step)

$$\phi_j = \frac{1}{m}\sum_{i=1}^{m} \omega_j^{(i)}, \tag{4.26}$$

$$\mu_j = \frac{\sum_{i=1}^{m} \omega_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} \omega_j^{(i)}}, \tag{4.27}$$

$$\Sigma_j = \frac{\sum_{i=1}^{m} \omega_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} \omega_j(i)} \tag{4.28}$$

4.6 Classification Error Method

The study includes four clustering methods (Adaboost Method, Genetic Programming, Maximum- Likelihood Method and Expectation-Maximization Method for defining LOS criteria of roads in urban Indian context. The best clustering method is determined by classification error process. First the data point are clustered by using clustering algorithm and after clustering , the classification error can be determined to get the most suitable clustering algorithm which is most appropriate for Indian context.

In the classification error process two sets of error were determined i.e. train set errors and test set errors. The total mean errors was determined to get the most suitable clustering algorithm.

Train Set Errors (Fraction of mistakes made on the training set).

Test Set Errors (Fraction of mistakes made on the testing set).

$$\begin{aligned} &Generalizationerror = pr_{x \in D}[h(x) \neq c(x)] \\ &= E[Testerror] \\ &= err(h) \end{aligned} \tag{4.29}$$

Where $D =$ Distribution factor.

$x =$ Data sets i.e. from $x_1, x_2, x_3, ...x_n$

$c(x) =$ Target function

$h(x) =$ Prediction value

$err(h) =$ Expected test error

<u>Classification Error Algorithm</u>:

In h error algorithm, two clusters are combined, when they have the same true mean. Consider the hypothesis $H_0 : \theta_i = \theta_j$, i.e. the true means of clusters c_i and c_j `are same. In other words combine $c_i$` and c_j, if H_0 is true.

For a fixed i, j it is easy to show that the statistic

$$d_{ij} = (\hat{x}_i - \hat{x}_j)^t [\hat{\sum}_i + \sum_j]^{-1} (\hat{x}_i - \hat{x}_j) \tag{4.30}$$

Follows a chi-square distribution with p degrees of freedom. If we denote the cumulative distribution function of a chi-square distribution with p degrees of freedom at a point t by $x_p(t)$, then $1 - x_p(d_{ij})$ gives the p value for accepting the hypothesis.

At 95% confidence, the merging of clusters stop, when minimum d_{ij} is greater than $x_p^{-1}(0.95)$.

Step-1: Input :-$(x_i, \sum_i), i = 1,2,3......n$ (4.31)

Step-2: Output:-Clusters (i), $i = 1,2,3,.........G$

For $i = 1$ to n do

Cluster $(i) = \{i\}$ (4.32)

Num cluster $= n$

Step-3: Loop

For $1 \le i < j \le$ Num clusters do

Calculate $d_{ij} = dist(cluster(i), cluster(j))$

When $d_{ij} = (\hat{x}_i - \hat{x}_j)^t [\hat{\sum}_i + \hat{\sum}_j]^{-1} (\hat{x}_i - \hat{x}_j)$ (4.33)

$(I, J) = \arg\ \min\ ij\ \ d_{ij}$

Step-4: if $d_{IJ} > x_p^{-1}(0.95)$ then

Break

$$
\begin{aligned}
&cluster(I) = cluster(I) \cup cluster(J)\\
&cluster(J) = cluster(Numclust)\\
&Numclust = Numclust - 1
\end{aligned}
\tag{4.34}
$$

Step-5: cluster $(i), i = 1,2,3,.............,G$ (4.35)

4.7 Cluster Validation Measures

Quality of clustering result incurred from a clustering algorithm can be checked by various cluster validation measures. These validation parameters have principally employed to appraise and compare whole partitions, ensuing from different algorithms or ensuing from the same algorithms under different parameters. Most common diligence of cluster validation is to ascertain the optimal number of cluster for a particular data set. (Bensaid et al; 1996) and as well clustering validation refers to procedure that evaluate the results of a clustering in a quantitative and objective fashion (Jain & Dubes, 1988).

Steps of cluster validation process:

Step-1:-Determining the clustering tendency of a set of data, i.e. distinguishing whether non-random structure actually exists in the data.

Step-2:-Comparing the results of a cluster analysis to externally known results, e.g.:- To externally given class levels.

Step-3:-Evaluating how well the results of a cluster analysis fit the data without reference to external information.

Step-4:-Comparing the results of two different sets of cluster analyses to determine which is better.

Step-5:-Determine the number of clusters.

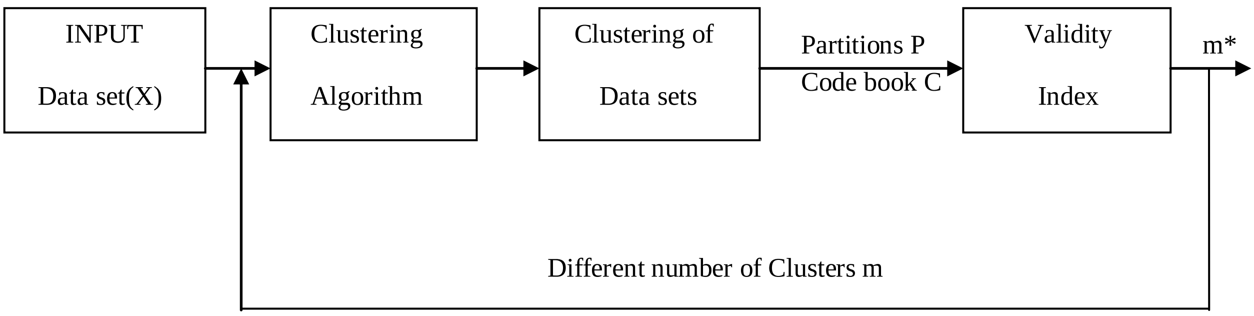


Figure-4.5: Flow Chart of Cluster Validation Measures

Primarily four parameters such as compactness, separability, exclusiveness and incorrectness have to be employed to delineate the objective function of cluster validity. The objective function is a combination of versatile four parameters.

The objective function is defined for a data set of 'n' number of dimensions. The verbal function, on considering a single cluster, is defined as:-

Objective function (OBF) = Min (compactness) +Max (separability) +Max (exclusiveness) +Min (incorrectness)

$$\text{OBF} = \{Min[c = \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n}\left\|x_{ij}-y_{ij}\right\|^2] + Max\,[s = \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n}\left\|c_i - c_j\right\|^2] +$$

$$Max[\frac{1}{(2\Pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{1}{2}(x-\mu)'}\Sigma^{-1}(x-\mu)] + Min[I = E(L(x,\mu))]\} \quad (4.36)$$

Where

$$x_i \in X_i[i = 1,2,3,.....,N)$$

$$y_i \in Y_i[i = 1,2,3,.....,N) \quad (4.37)$$

x_{ij}, y_{ij} = Data point co-ordinates of the i^{th} cluster in the j^{th} dimension.

$\mu = (\mu_1, \mu_2, \mu_3, ..., \mu_n)$

$k =$ Total number of dimensions of a dataset.

$E(\ ) =$ Expected value of a function.

$c =$ Compactness measure.

$s =$ Separability measure.

$Ex =$ Exclusiveness measure.

$I =$ In-correctness measure.

$n =$ Total number of data items in the i^{th} cluster $[i = 1,2,3,.....N]$

$N =$ Total number of clusters.

$c_i, c_j =$ Computer centroid values of cluster i and j.

$\sigma, \Sigma =$ Variance value of the entire data set.

4.8 Validity Index

Different validity indices have been suggested for the study, but none of them is blemish by oneself, and consequently various indices have employed in this study, such as:- Homogeneity-separation index(HSI), Rand index(RI), Adjusted rand index(ARI), Hubert index(HI) and Mirkin index(MI).

A) Homogeneity-separation index (HSI)

The Index was suggested by Shamir and Sharan(2002). Homogeneity is calculated as the average distance between each gene expression profile and the canter of the cluster it belongs to. That is,

$$H_{ave} = \frac{1}{N_{gene}} \sum_i D(g_i, C(g_i)) \tag{4.38}$$

Where g_i is the ith gene and $C(g_i)$ is the center of the cluster that g_i belongs to; N_{gene} is the total number of genes; D is the distance function.

Separation is calculated as the weighted average distance between cluster canters:

$$S_{ave} = \frac{1}{\sum_{i \neq j} N_{ci} N_{cj}} \sum_{i \neq j} N_{ci} N_{cj} D(C_i, C_j) \tag{4.39}$$

Where C_i and C_j are the canters of ith and jth clusters, and N_{ci} and N_{cj} are the number of genes in the ith and jth clusters. Thus H_{ave} reflects the compactness of the clusters while S_{ave} reflects the overall distance between clusters. Decreasing H_{ave} or increasing S_{ave} suggests an improvement in the clustering results.

B) Rand index (RI)

Rand Index is the fraction of agreements with respect to element pairs that are either clustered together in both clustering's or clustered apart in both clusterings.

$$RI = \frac{N_{11} + N_{00}}{N_{11} + N_{00} + N_{01} + N_{10}} \tag{4.40}$$

Where N_{11} = number of pairs of points clustered together in both clusterings.

N_{00} = number of element pairs that both clusterings did not cluster together.

N_{01} = number of pairs clustered in second but not first clustering.

N_{10} = number of pairs clustered in first but not second clustering.

C) Adjusted rand index (ARI)

The Rand Index has been adjusted such that the normalized index has expected value 0 and value cannot exceed 1 (Meila 2007).

$$ARI = \frac{RI - E[R]}{1 - E[R]} \tag{4.41}$$

D) Hubert index(HI)

Hubert Γ is defined (Halkidi, Batistakis, & Vazirgiannis, 2002) as

$$\tau = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{k=i+1}^{N} d_{ik} Cl_{ik} \tag{4.42}$$

Where d_{ik} = distance between elements i and k.

Cl_{ik} = distance between clusters to which elements i and k belong (represented by centroids.)

Entropy:

Assuming that a point has equal probability of belonging to any cluster, the entropy of a clustering is defined as (Meila, 2007):

$$H(C) = -\sum_{i=1}^{k} p(i)\log p(i) \tag{4.43}$$

$$\text{Where } \; p(i) = \frac{n_i}{n} \tag{4.44}$$

k = number of clusters.

E) Mirkin index(MI)

The Mirkin index which is also known as Equivalence Mismatch Distance is defined by

$$M(C,C') = \sum_{i=1}^{k} \left|C_i\right|^2 + \sum_{j=1}^{l} \left|C_j\right|^2 - 2\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij}^{\;2} \tag{4.45}$$

It corresponds to the Hamming distance for binary vectors if the set of all pairs of elements is enumerated and a clustering is represented by a binary vector defined on this enumeration. An advantage is the fact that this distance is a metric on $p(x)$. However, this measure is very sensitive to cluster sizes such that two clusterings that are ”at right angles” to each other (i.e. each cluster in one clustering contains the same amount of elements of each of the clusters of the other clustering) are closer to each other than two clusterings for which one is a refinement of the other.

Chapter 5

Result and Analysis

5.1 Introduction

Result of cluster analysis is discussed in this chapter. In clustering, the Adaboost, Genetic Programming, Maximum Likelihood Method and Expectation-Maximization algorithms were used. A basic idea about the algorithm used is described. By clustering using the above four algorithm the urban street segments were classified into four classes from the free flow speed data. After defining the segment into a particular class of urban street speed range for six LOS categories were found out.

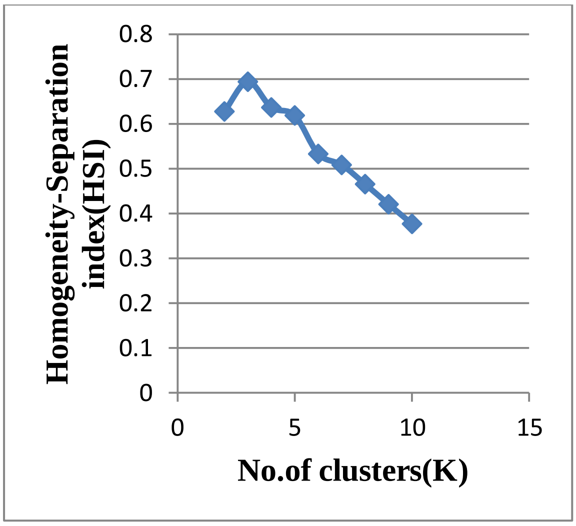
5.2 Application of Cluster Analysis Methods in Defining LOS Criteria of Urban Streets.

Average travel speeds were calculated direction wise on each segment. Four advanced cluster analysis techniques (Adaboost, GP, ML & EM) were applied in two stages. Firstly, clustering methods were applied on average free flow speeds of all segments and free-flow speeds were classified into four groups. Each range of free flow speed found out indicates to an urban street class of I to IV. Secondly, clustering methods were applied on average travel speeds that were collected during peak and off peak hours on street segments for each of the urban street classes. In the second case, speeds were classified into six groups (A to F) for six categories of levels of service; thus speed ranges for level of service categories were defined in Indian context.

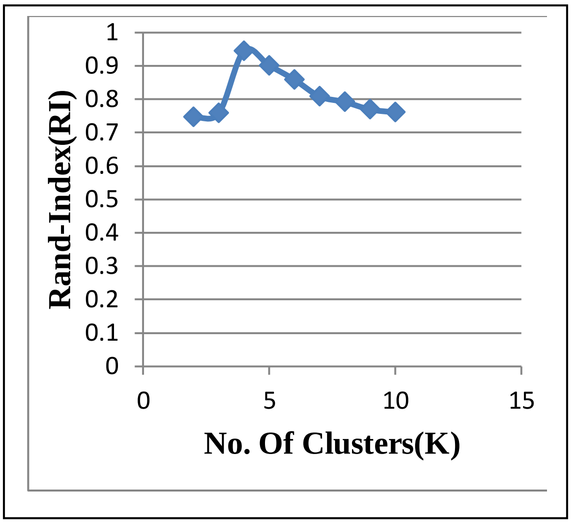
5.2.1 Adaboost Method

The free flow speed data acquired through GPS receiver was clustered using the Adaboost Algorithm. For determination of the parametric value of validation measures, free flow speed data were used. In this research five validation parameters were used. Value of validation parameters were obtained for 2 to 10 number of cluster and were plotted in Figure 5.1 (A) to Figure 5.1 (E).

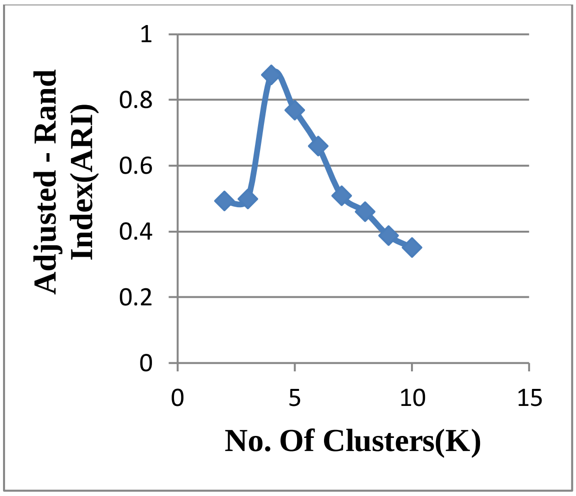
These Five number of validation parameters were used to know the optimum number of cluster for this particular data set of free flow speed. By knowing the optimum number of cluster we can classify the urban street segments into that number of Urban street classes. It is always considered that lesser number of clusters is better if variation in validation parameters is minimal. Literature says that the highest value of Homogeneity-Separation Index (HIS) signifies the optimal number of cluster for a particular set of data. Figure 5.1 (A) shows that the index is highest for 3 number of clusters. Also, available literature says that the highest value of Rand-Index (RI), Adjusted Rand Index(ARI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 5.1 (B)&(C). For Mirkin Index (MI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes Upward.. Figure 5.1(D) shows the Mirkin Index (MI). The Hubert Index (HI) gives that the highest value is the optimum number of clusters. Figure 5.1(E) describes the highest value is the optimum number of clusters. Out of five validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the Adaboost Algorithm.



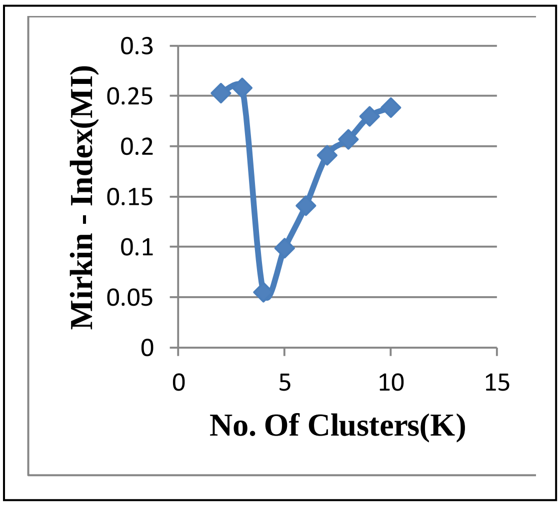
A: HSI vs. Number of cluster



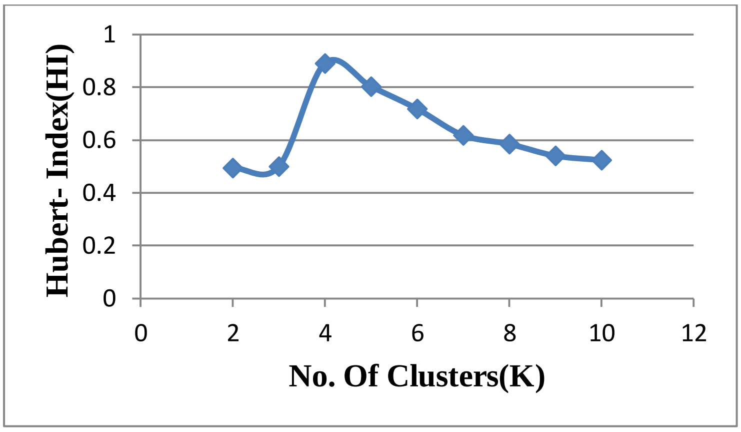
B: RI vs. Number Of Cluster



C: ARI vs. Number Of Cluster



D: MI vs. Number Of Cluster



E: HI vs. Number Of Cluster

Figure-5.1: Validation measures for optimal number of clusters using Adaboost Method

Figure 5.2 shows the speed ranges for different urban street classes. Different symbol in the plot used for different urban street class. It is observed from the collected data set that when a street segment falls under particular urban street class is agreed with the geometric and surrounding environmental condition of the road segments as well.

It has been found that there is very good correlation between free flow speed and geometric and environmental characteristics of streets under considerations.

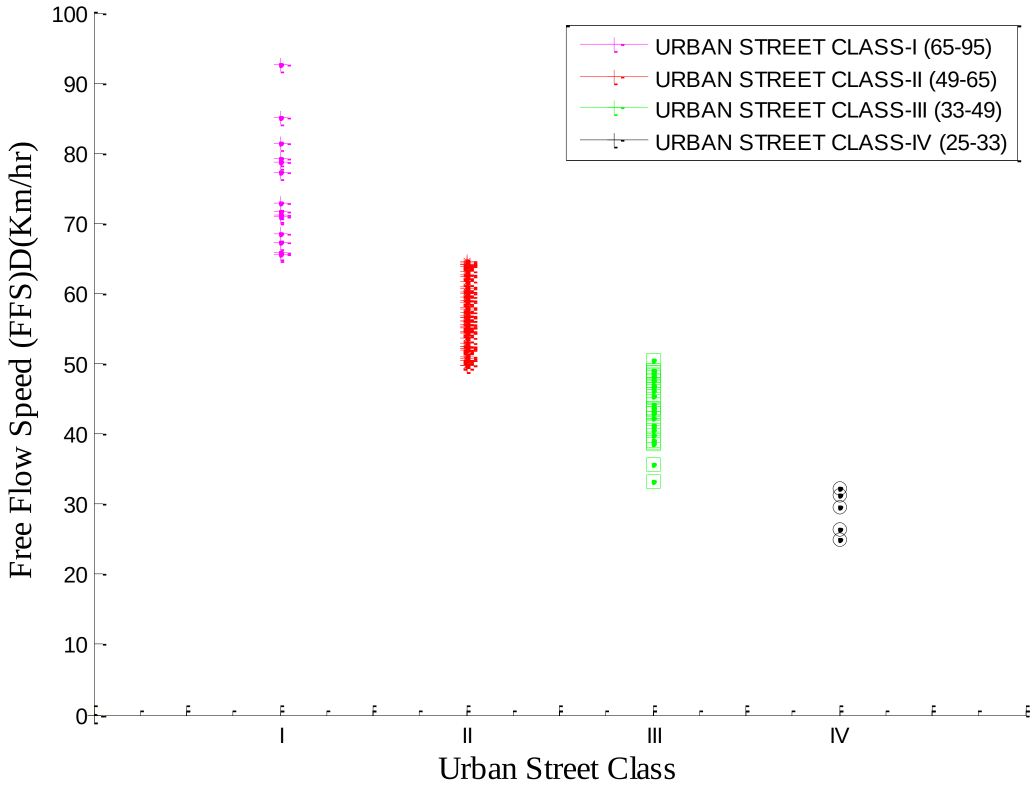
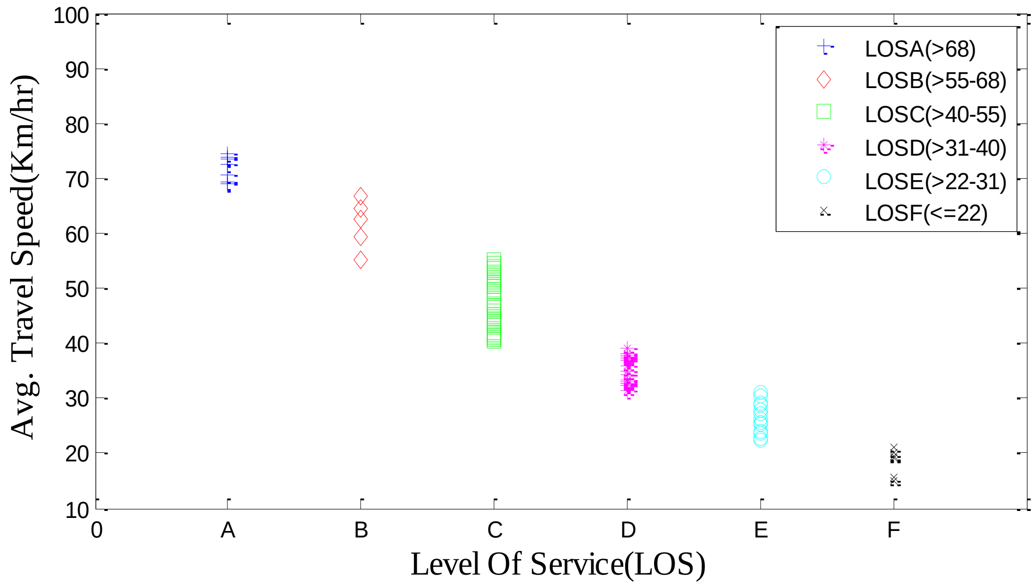
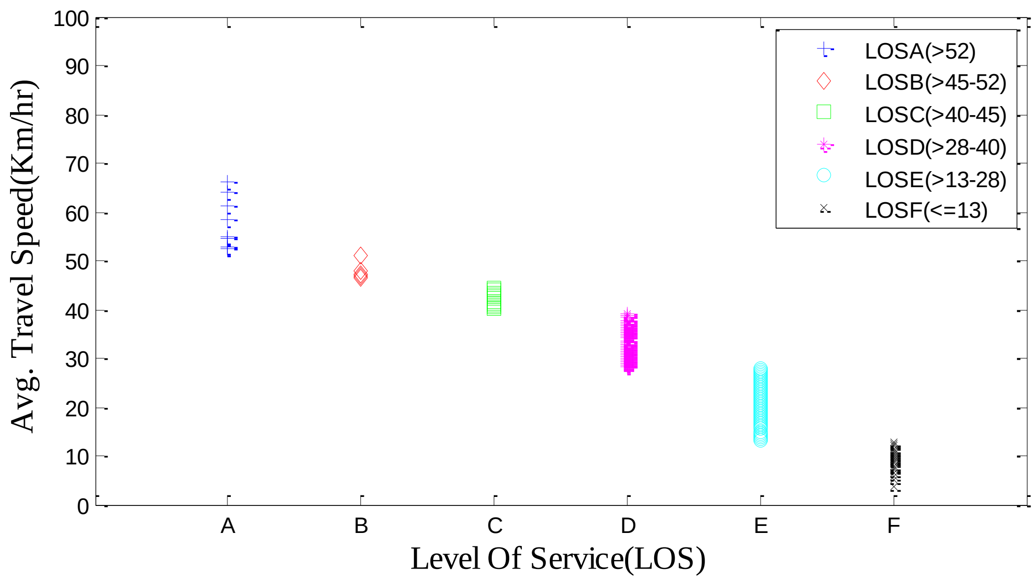


Figure-5.2: ADABOOST Clustering of FFS for Urban Street Classification

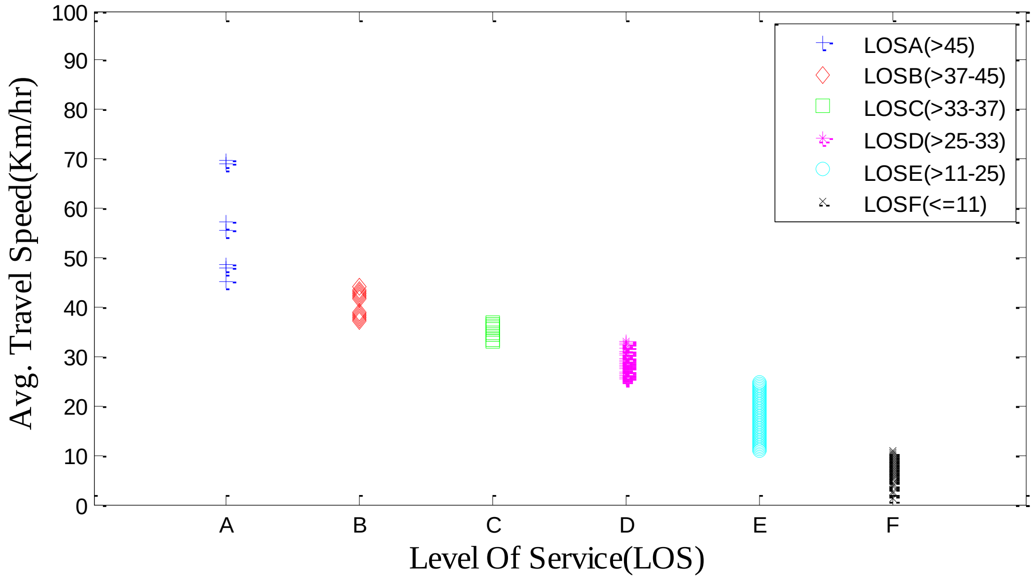
After classification of urban streets into number of classes, direction wise average travel speed on street segments during both peak and off peak hours were clustered using Adaboost Algorithm to find the speed range of level of service categories. In fig. 5.3 the speed values are shown by different symbols depending on to which LOS category they belong. The legends in fig.5. 3 (A-D) gives the speed ranges for the six LOS categories obtained by using Adaboost clustering. The speed ranges for LOS categories found using Adaboost clustering is also shown in Table 5.1.



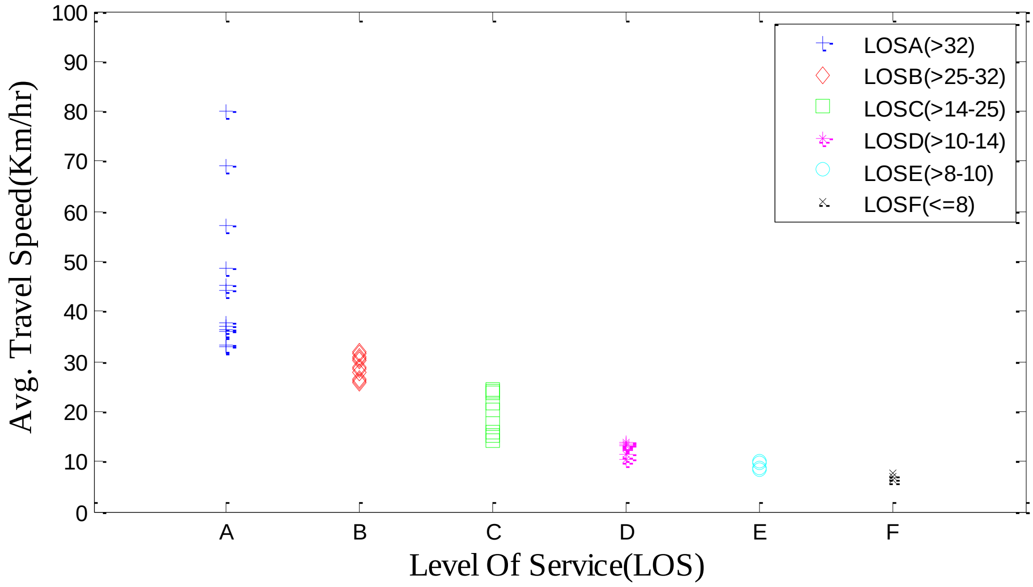
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure-5.3: Level of service of urban street classes (I-IV) using Adaboost clustering on average travel speeds

Table-5.1: Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by ADABOOST Method

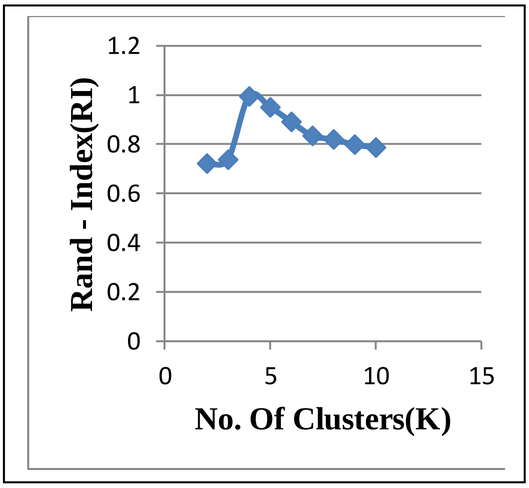
Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	65 to 95 km/h	49 to 65 km/h	33 to 49 km/h	25 to 33 km/h
Typical FFS	72km/h	58km/h	39km/h	27 km/h
LOS	Average Travel Speed (Km/h)			
A	>68	>52	>45	>32
B	>55-68	>45-52	>37-45	>25-32
C	>40-55	>40-45	>33-37	>14-25
D	>31-40	>28-40	>25-33	>10-14
E	>22-31	>13-28	>11-25	>8-10
F	≤22	≤13	≤11	≤8

5.2.2 Genetic Programming

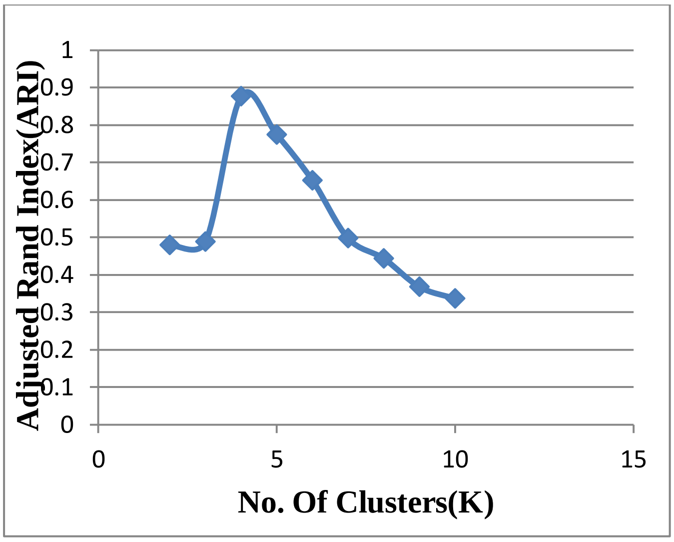
The free flow speed data acquired through GPS receiver was clustered using the Genetic Programming. For determination of the parametric value of validation measures, free flow speed data were used. In this research five validation parameters were used. Value of validation parameters were obtained for 2 to 10 number of cluster and were plotted in Figure 5.4 (A) to Figure 5.4 (E).

These Five number of validation parameters were used to know the optimum number of clusters for this particular data set of free flow speed. By knowing the optimum number of clusters we can classify the urban street segments into that number of Urban street classes. It is always considered that lesser number of clusters is better if variation in validation parameters is minimal. Literature says that the highest value of Homogeneity-Separation Index (HIS) signifies the optimal number of cluster for a particular set of data. Figure 5.4 (A) shows that the index is highest for 3 number of clusters. Also, available literature says that the highest value of Rand-Index (RI), Adjusted Rand Index (ARI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 5.4 (B)&(C). ForMirkin Index (MI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes Upward.. Figure 5.4(D) shows the Mirkin Index (MI). The Hubert Index(HI) gives that the highest value is the optimum number of clusters. Figure 5.4(E) describes the highest value is the optimum number of clusters. Out of five validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the Genetic Programming.

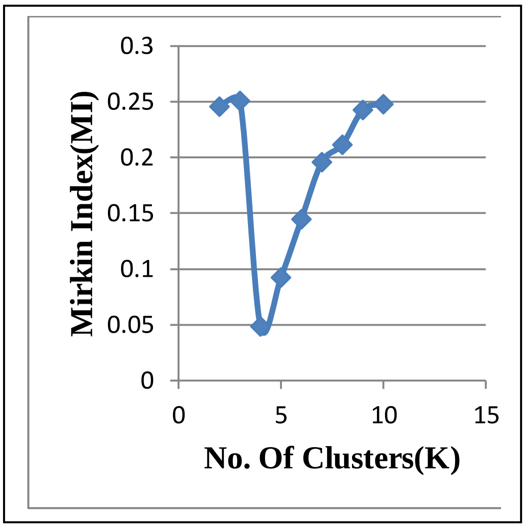
A: HSI vs. Number of cluster



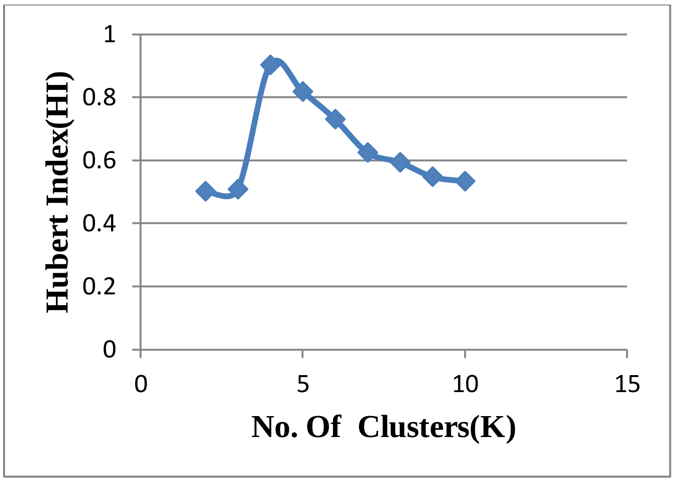
B: RI vs. Number Of Cluster



C: ARI vs. Number Of Cluster



D: MI vs. Number Of Cluster



E: HI vs. Number Of Cluster

Figure-5.4: Validation measures for optimal number of clusters using Genetic Programming

Figure 5.5 shows the speed ranges for different urban street classes. Different symbol in the plot used for different urban street class. It is observed from the collected data set that when a street segment falls under particular urban street class is agreed with the geometric and surrounding environmental condition of the road segments as well.

It has been found that there is very good correlation between free flow speed and geometric and environmental characteristics of streets under considerations.

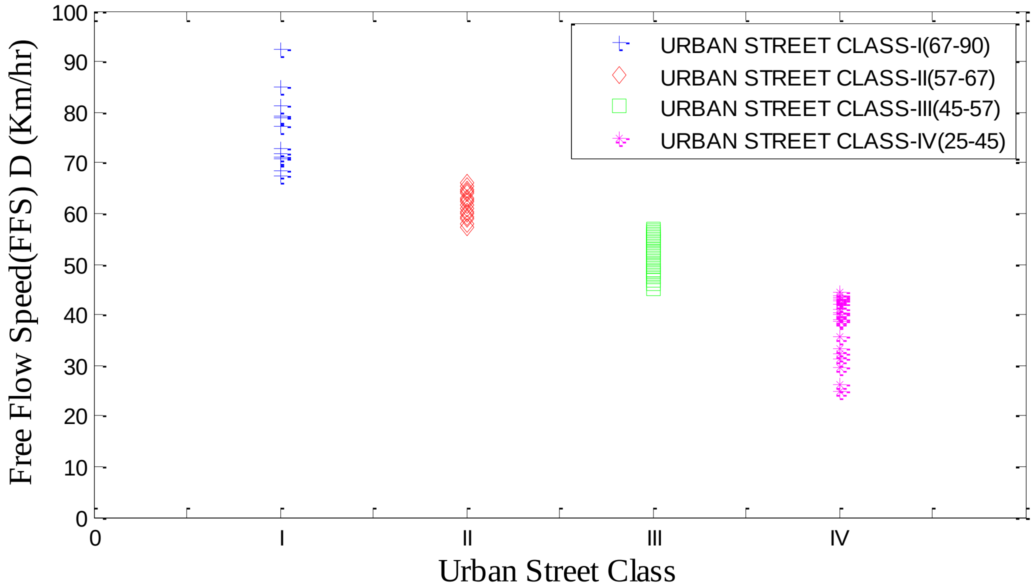
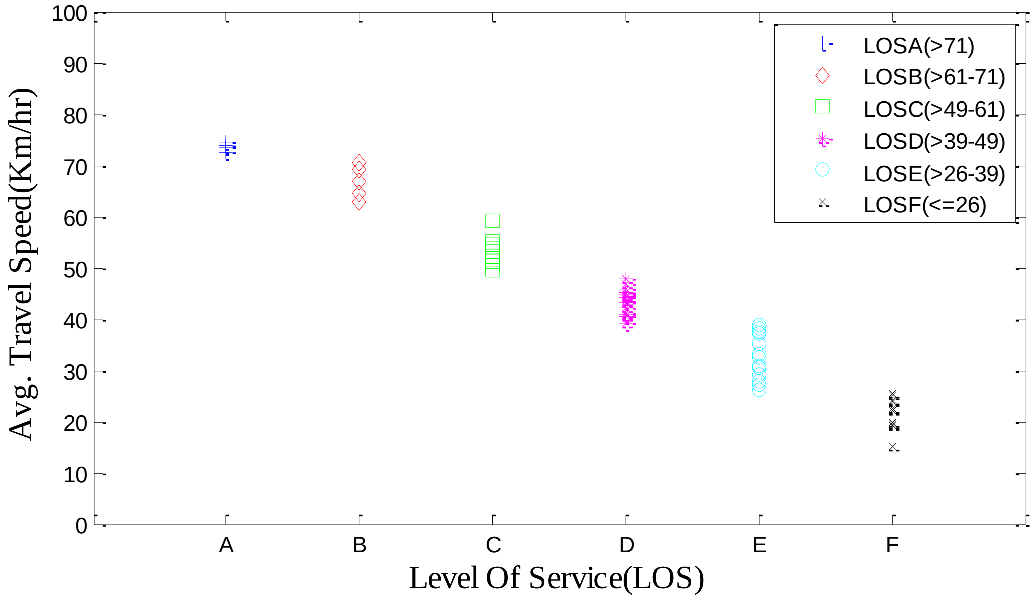
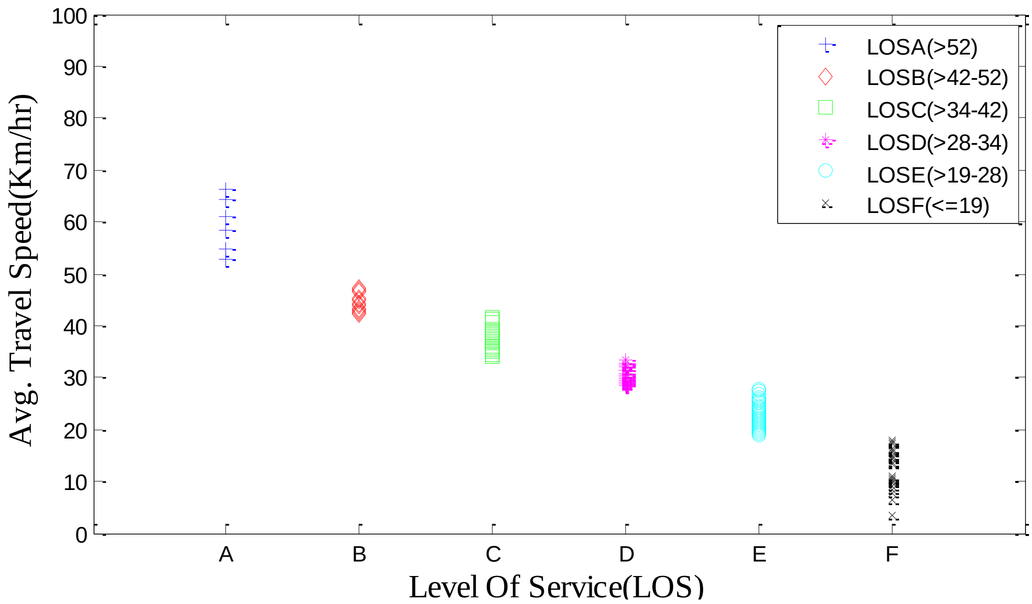


Figure-5.5: GP Clustering of FFS for Urban Street Classification

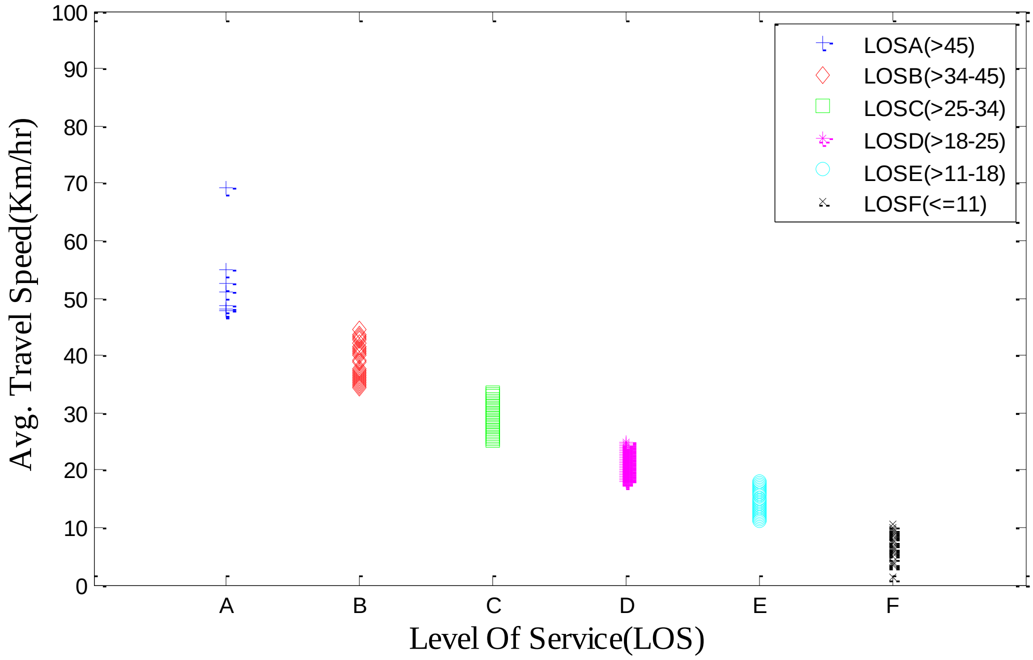
After classification of urban streets into number of classes, direction wise average travel speed on street segments during both peak and off peak hours were clustered using GP to find the speed range of level of service categories. In fig. 5.6 the speed values are shown by different symbols depending on to which LOS category they belong. The legends in fig.5. 6 (A-D) gives the speed ranges for the six LOS categories obtained by using GP. The speed ranges for LOS categories found using GP is also shown in Table 5.2.



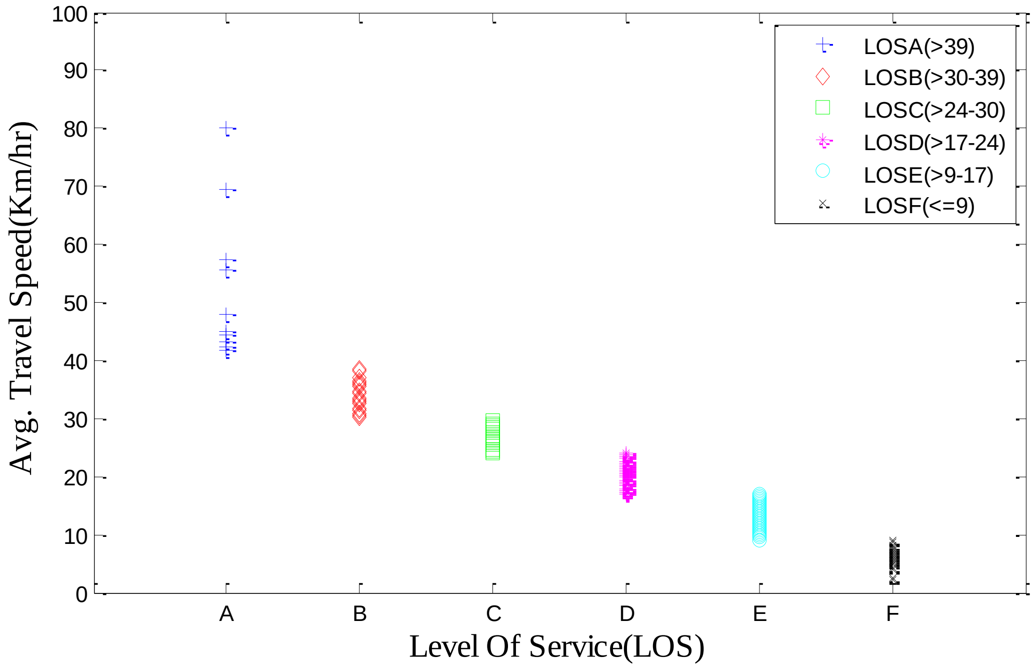
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure-5.6: Level of service of urban street classes (I-IV) using Genetic Programming on average travel speeds

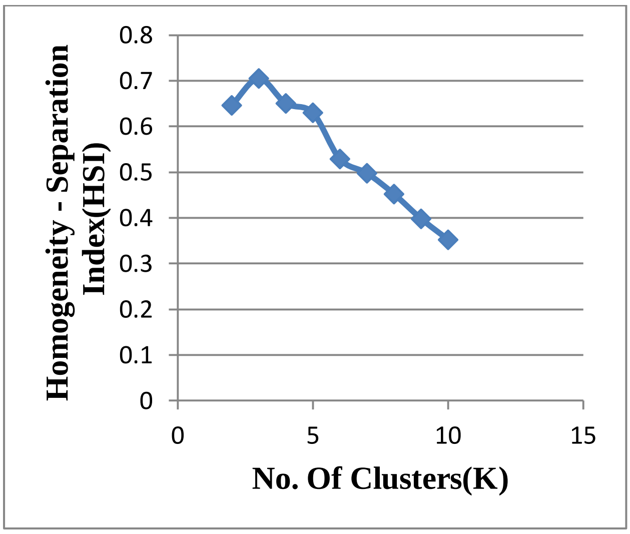
Table-5.2: Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by GP Method

Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	67to 90 km/h	57to 67 km/h	45 to 57 km/h	25 to 45 km/h
Typical FFS	72km/h	61km/h	51km/h	37 km/h
LOS	Average Travel Speed (Km/h)			
A	>71	>52	>45	>39
B	>61-71	>42-52	>34-45	>30-39
C	>49-61	>34-42	>25-34	>24-30
D	>39-49	>28-34	>18-25	>17-24
E	>26-39	>19-28	>11-18	>9-17
F	≤26	≤19	≤11	≤9

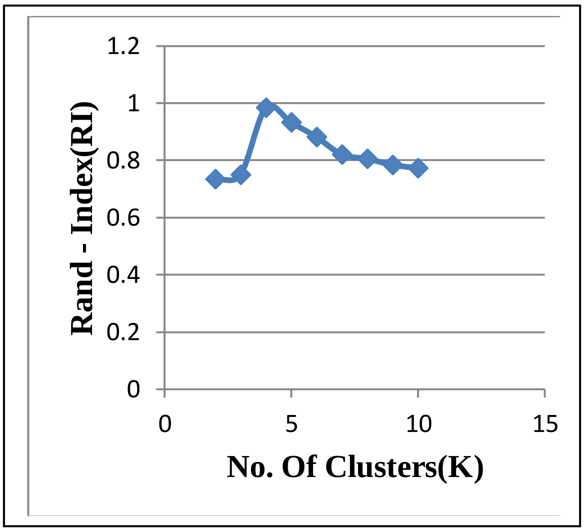
5.2.3 Maximum Likelihood Method

The free flow speed data acquired through GPS receiver was clustered using the Maximum Likelihood Method. For determination of the parametric value of validation measures, free flow speed data were used. In this research five validation parameters were used. Value of validation parameters were obtained for 2 to 10 number of cluster and were plotted in Figure 5.7 (A) to Figure 5.7 (E).

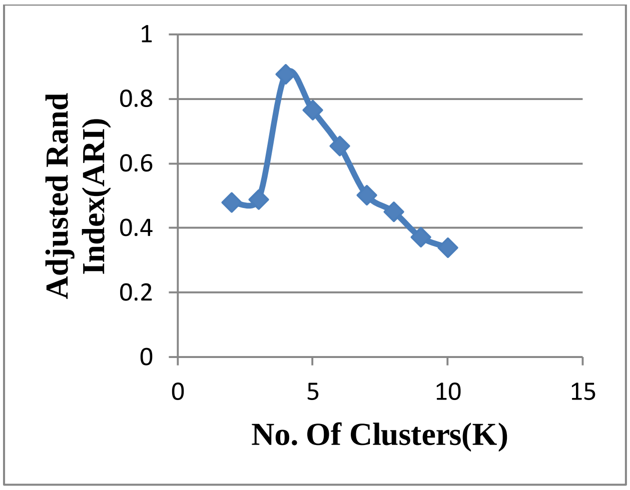
These Five number of validation parameters were used to know the optimum number of cluster for this particular data set of free flow speed. By knowing the optimum number of cluster we can classify the urban street segments into that number of Urban street classes. It is always considered that lesser number of clusters is better if variation in validation parameters is minimal. Literature says that the highest value of Homogeneity-Separation Index (HIS) signifies the optimal number of cluster for a particular set of data. Figure 5.7 (A) shows that the index are highest for 3 number of clusters. Also, available literature says that the highest value of Rand-Index (RI), Adjusted Rand Index (ARI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 5.7 (B)&(C). For Mirkin Index (MI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes Upward.. Figure 5.7(D) shows the Mirkin Index (MI). The Hubert Index (HI) gives that the highest value is the optimum number of clusters. Figure .7(E) describes the highest value is the optimum number of clusters. Out of five validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the Maximum Likelihood Method.



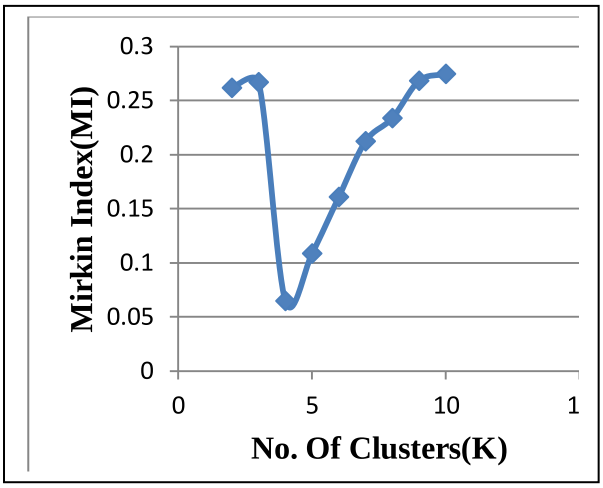
A: HSI vs. Number of cluster



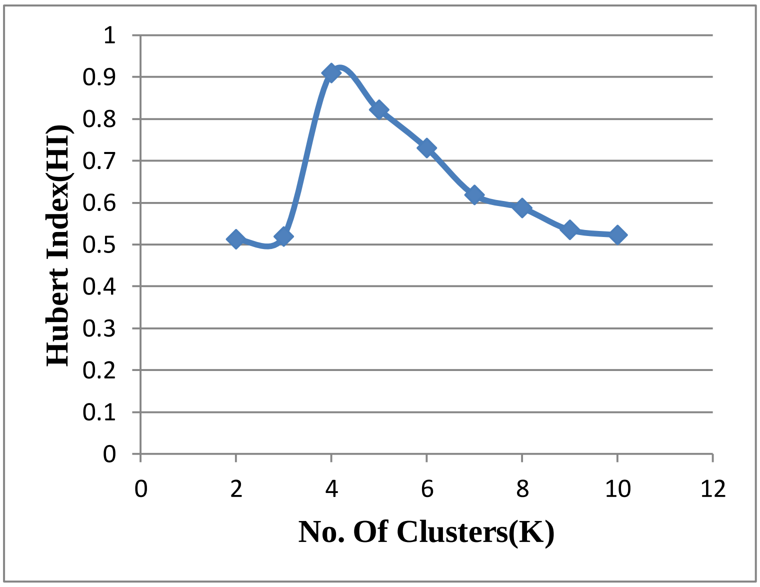
B: RI vs. Number Of Cluster



C: ARI vs. Number Of Cluster



D: MI vs. Number Of Cluster



E: HI vs. Number Of Cluster

Figure-5.7: Validation measures for optimal number of clusters using Maximum Likelihood Method

Figure 5.8 shows the speed ranges for different urban street classes. Different symbol in the plot used for different urban street class. It is observed from the collected data set that when a street segment falls under particular urban street class is agreed with the geometric and surrounding environmental condition of the road segments as well.

It has been found that there is very good correlation between free flow speed and geometric and environmental characteristics of streets under considerations.

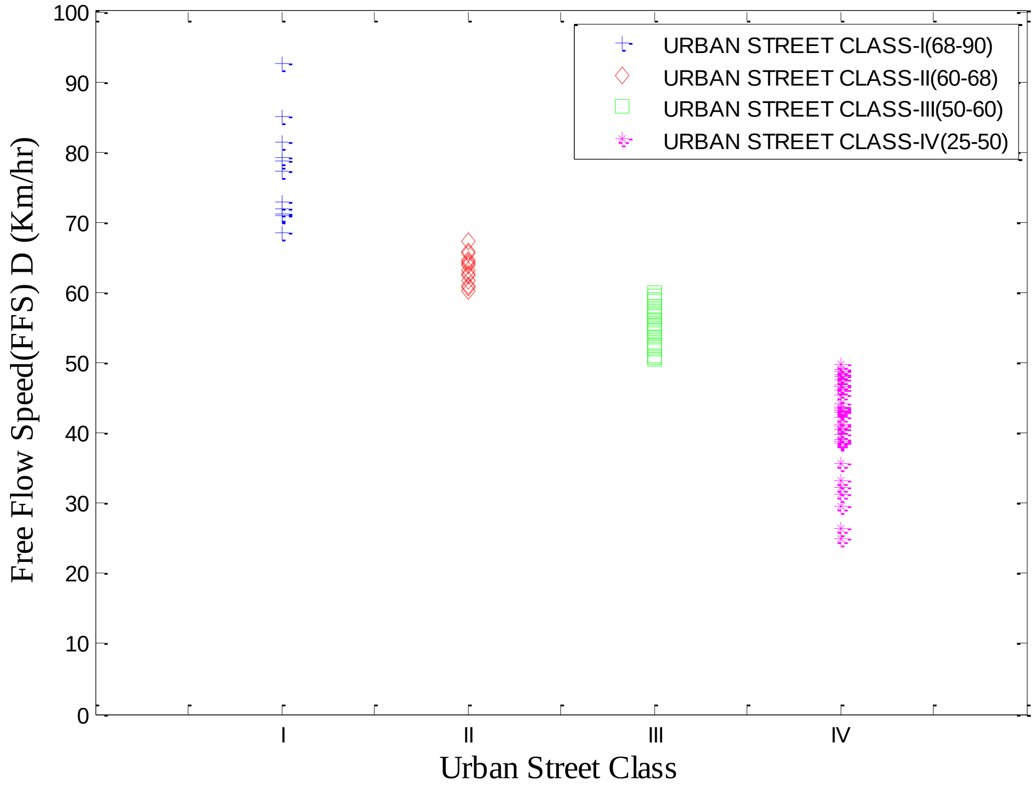
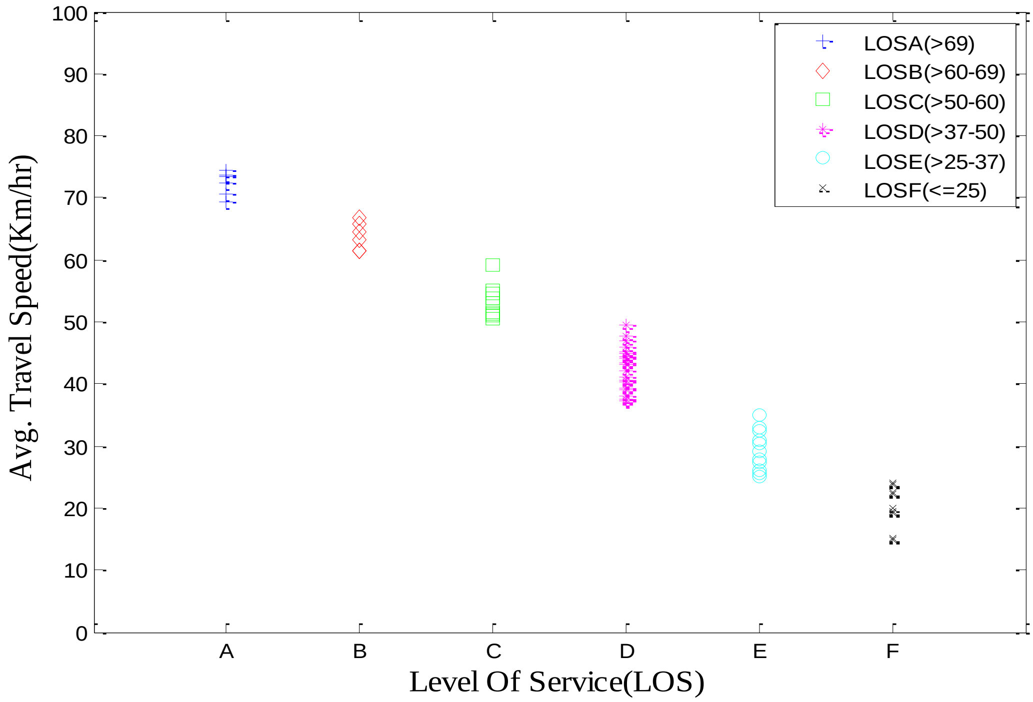


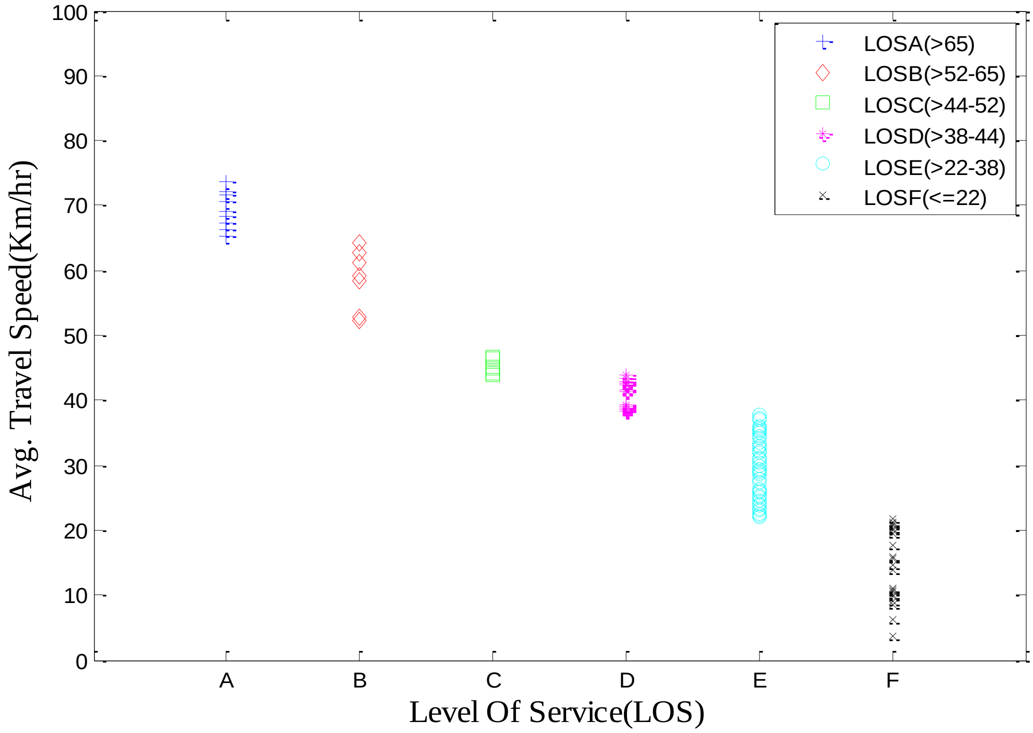
Figure-5.8: Maximum Likelihood Method of FFS for Urban Street Classification

After classification of urban streets into number of classes, direction wise average travel speed on street segments during both peak and off peak hours were clustered using Maximum

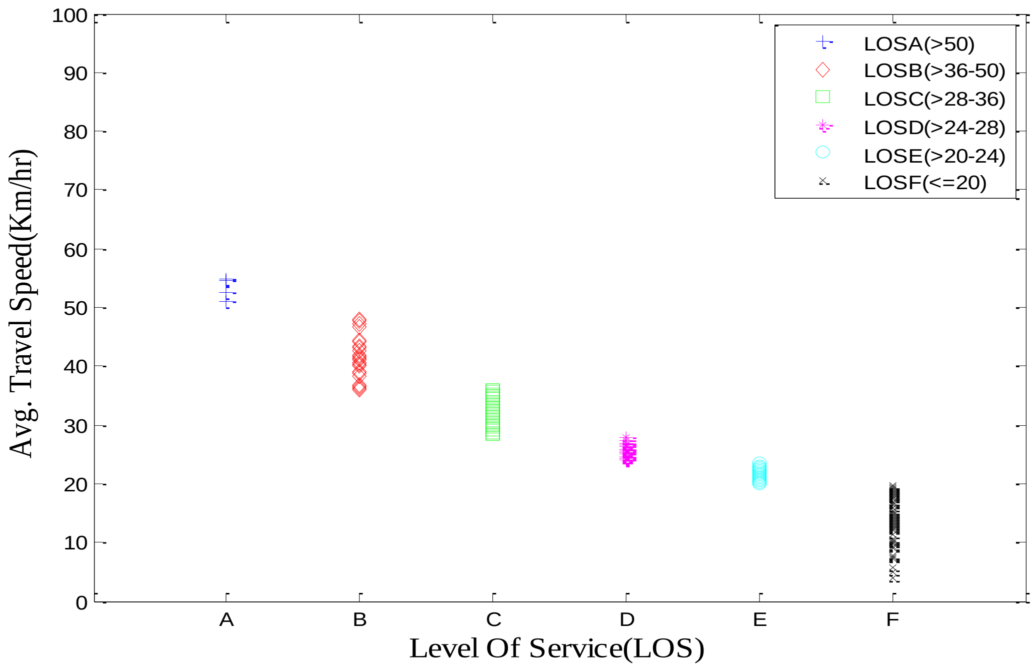
Likelihood Method to find the speed range of level of service categories. In fig. 5.9 the speed values are shown by different symbols depending on to which LOS category they belong. The legends in fig.5. 9(A-D) gives the speed ranges for the six LOS categories obtained by using Maximum Likelihood Method. The speed ranges for LOS categories found using Maximum Likelihood Method is also shown in Table 5.3.



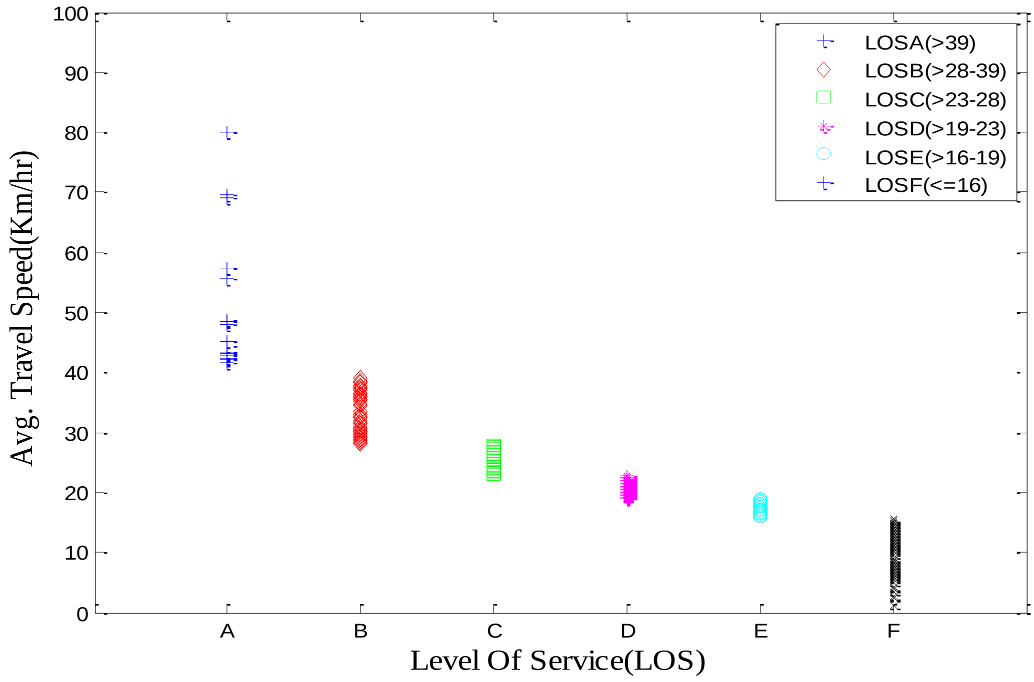
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure-5.9: Level of service of urban street classes (I-IV) using Maximum Likelihood Method on average travel speed

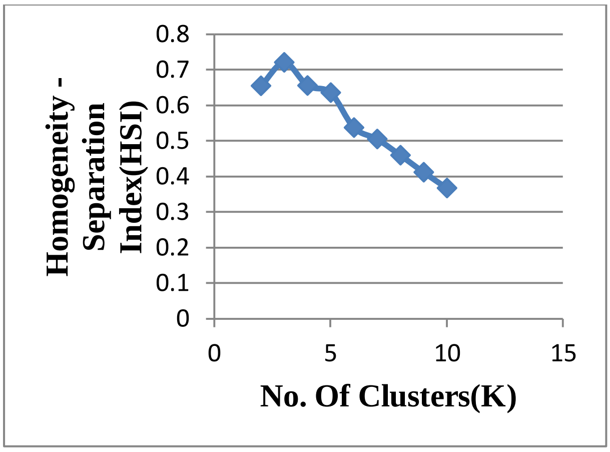
Table-5.3: Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by Maximum-Likelihood Method

Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	68to 90 km/h	60to 68 km/h	50 to 60 km/h	25 to 50 km/h
Typical FFS	75km/h	61km/h	57km/h	43 km/h
LOS	Average Travel Speed (Km/h)			
A	>69	>65	>50	>39
B	>60-69	>52-65	>36-50	>28-39
C	>50-60	>44-52	>28-36	>23-28
D	>37-50	>38-44	>24-28	>19-23
E	>25-37	>22-38	>20-24	>16-19
F	≤25	≤22	≤20	≤16

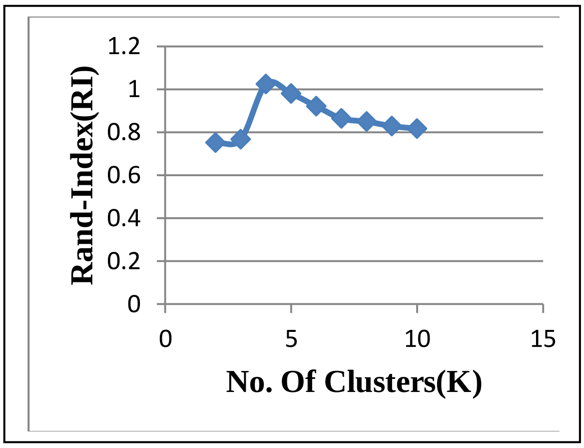
5.2.4 Expectation-Maximization Method

The free flow speed data acquired through GPS receiver was clustered using the Expectation-Maximization Method. For determination of the parametric value of validation measures, free flow speed data were used. In this research five validation parameters were used. Value of validation parameters were obtained for 2 to 10 number of cluster and were plotted in Figure 5.10 (A) to Figure 5.10 (E).

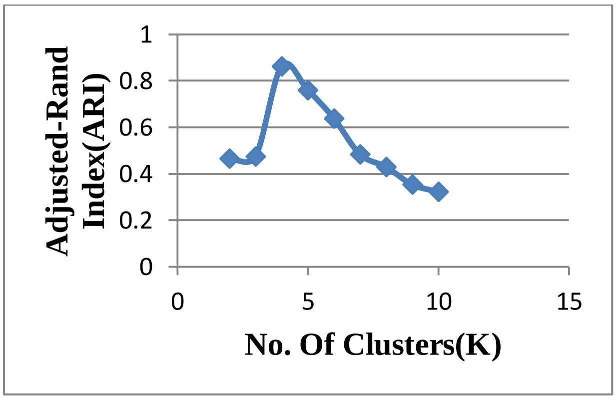
These Five number of validation parameters were used to know the optimum number of cluster for this particular data set of free flow speed. By knowing the optimum number of cluster we can classify the urban street segments into that number of Urban street classes. It is always considered that lesser number of clusters is better if variation in validation parameters is minimal. Literature says that the highest value of Homogeneity-Separation Index (HIS) signifies the optimal number of cluster for a particular set of data. Figure 5.10 (A) shows that the index are highest for 3 number of clusters. Also, available literature says that the highest value of Rand-Index (RI), Adjusted Rand Index (ARI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 5.10 (B)&(C). ForMirkin Index (MI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes Upward.. Figure 5.10(D) shows the Mirkin Index (MI). The Hubert Index (HI) gives that the highest value is the optimum number of clusters. Figure 5.10(E) describes the highest value is the optimum number of clusters. Out of five validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the Expectation-Maximization Method.



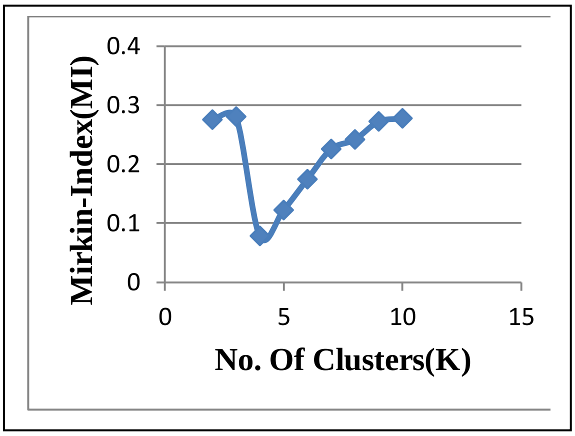
A: HSI vs. Number of cluster



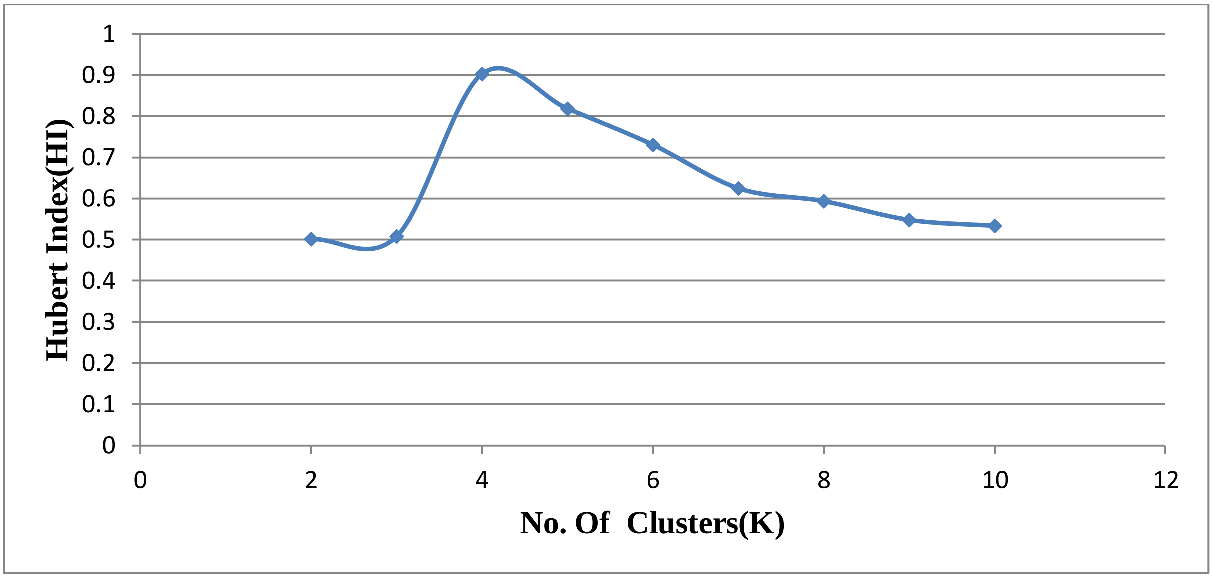
B: RI vs. Number Of Cluster



C: ARI vs. Number Of Cluster



D: MI vs. Number Of Cluster



E: HI vs. Number Of Cluster

Figure-5.10: Validation measures for optimal number of clusters using Expectation-Maximization Method

Figure 5.11 shows the speed ranges for different urban street classes. Different symbol in the plot used for different urban street class. It is observed from the collected data set that when a street segment falls under particular urban street class is agreed with the geometric and surrounding environmental condition of the road segments as well.

It has been found that there is very good correlation between free flow speed and geometric and environmental characteristics of streets under considerations.

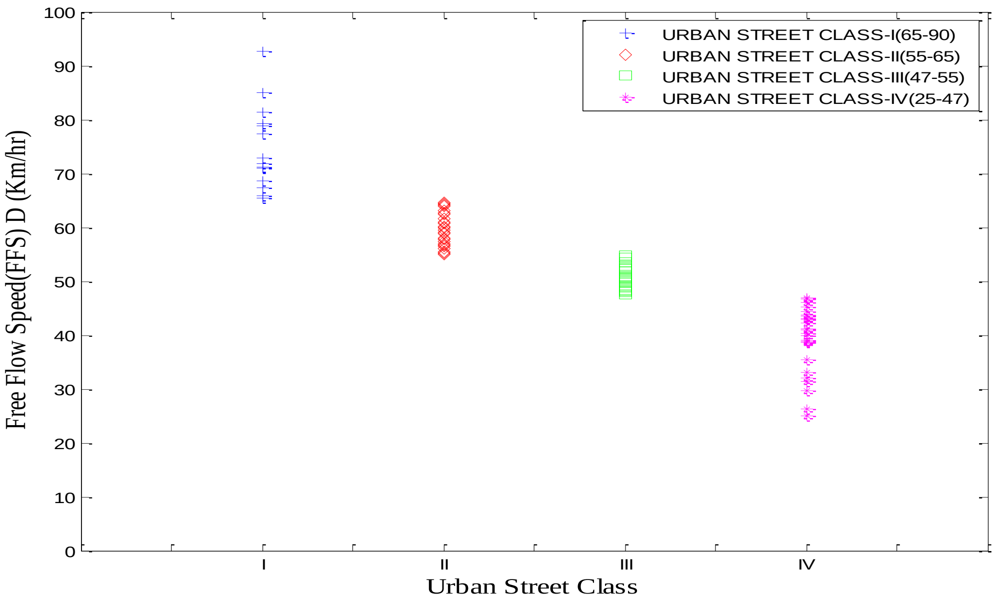
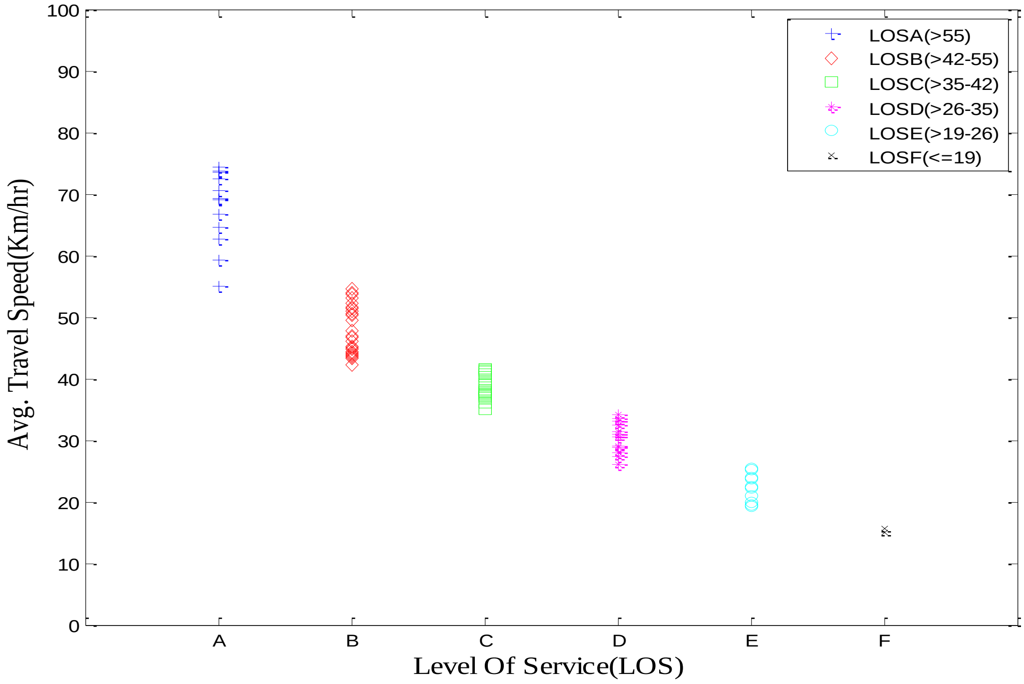
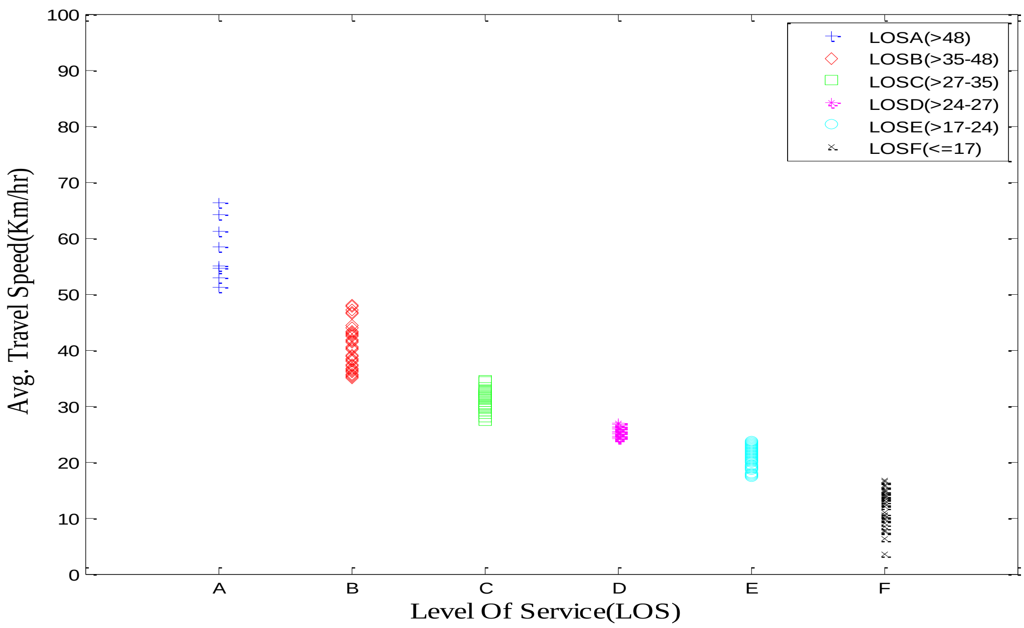


Figure -5.11: Expectation-Maximization Method of FFS for Urban Street Classification

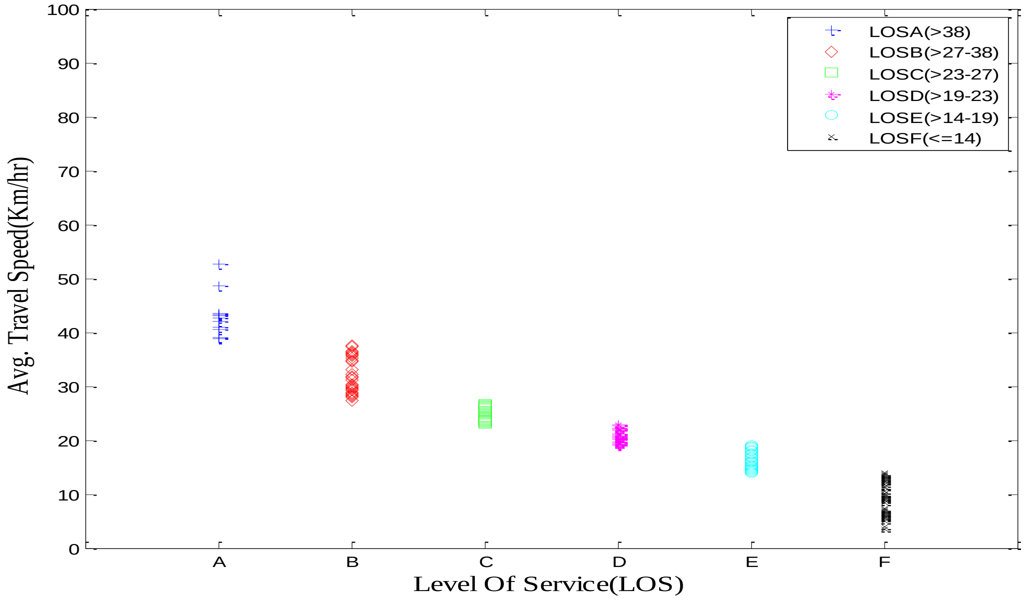
After classification of urban streets into number of classes, direction wise average travel speed on street segments during both peak and off peak hours were clustered using Expectation-Maximization Method to find the speed range of level of service categories. In fig. 5.12 the speed values are shown by different symbols depending on to which LOS category they belong. The legends in fig.5.12 (A-D) gives the speed ranges for the six LOS categories obtained by using Expectation-Maximization Method. The speed ranges for LOS categories found using Expectation-Maximization Method is also shown in Table 5.4



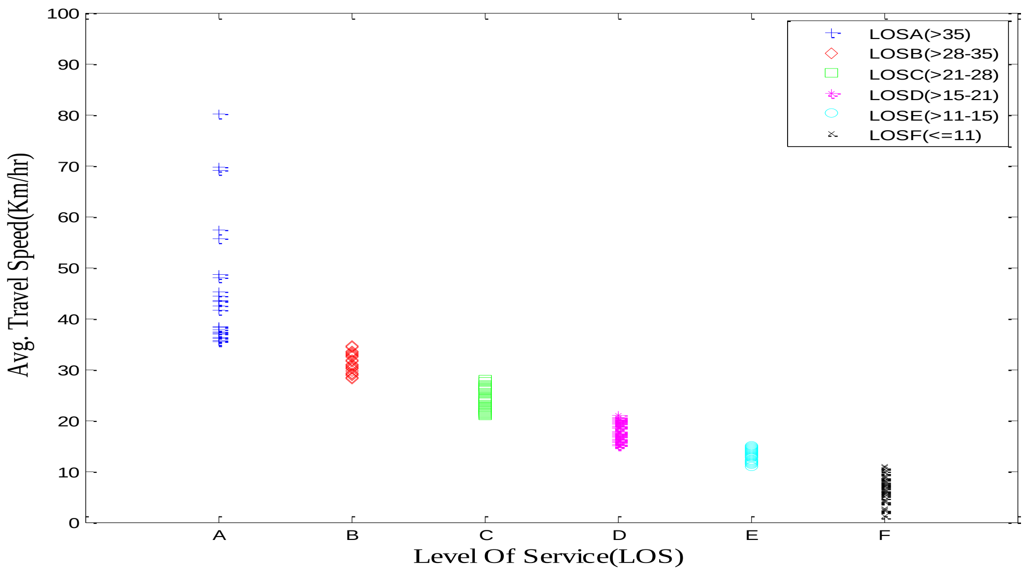
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure- 5.12: Level of service of urban street classes (I-IV) using Expectation-Maximization Method on average travel speed

Table-5.4: Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by EM Method

Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	65to 90 km/h	55to 65 km/h	47 to 55 km/h	25 to 47 km/h
Typical FFS	70km/h	60km/h	50km/h	35 km/h
LOS	Average Travel Speed (Km/h)			
A	>55	>48	>38	>35
B	>42-55	>35-48	>27-38	>28-35
C	>35-42	>27-35	>23-27	>21-28
D	>26-35	>24-27	>19-23	>15-21
E	>19-26	>17-24	>14-19	>11-15
F	≤19	≤17	≤14	≤11

5.3 Representation of Free Flow Speed in Radar diagram

5.3.1 Radar Diagram

A radar chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point (Georg von Mayr 1877). The relative position and angle of the axis is typically uninformative and also radar chart is a chart that consists of sequence of equi-angular spokes is called radii. With each spoke representing one of the variable.

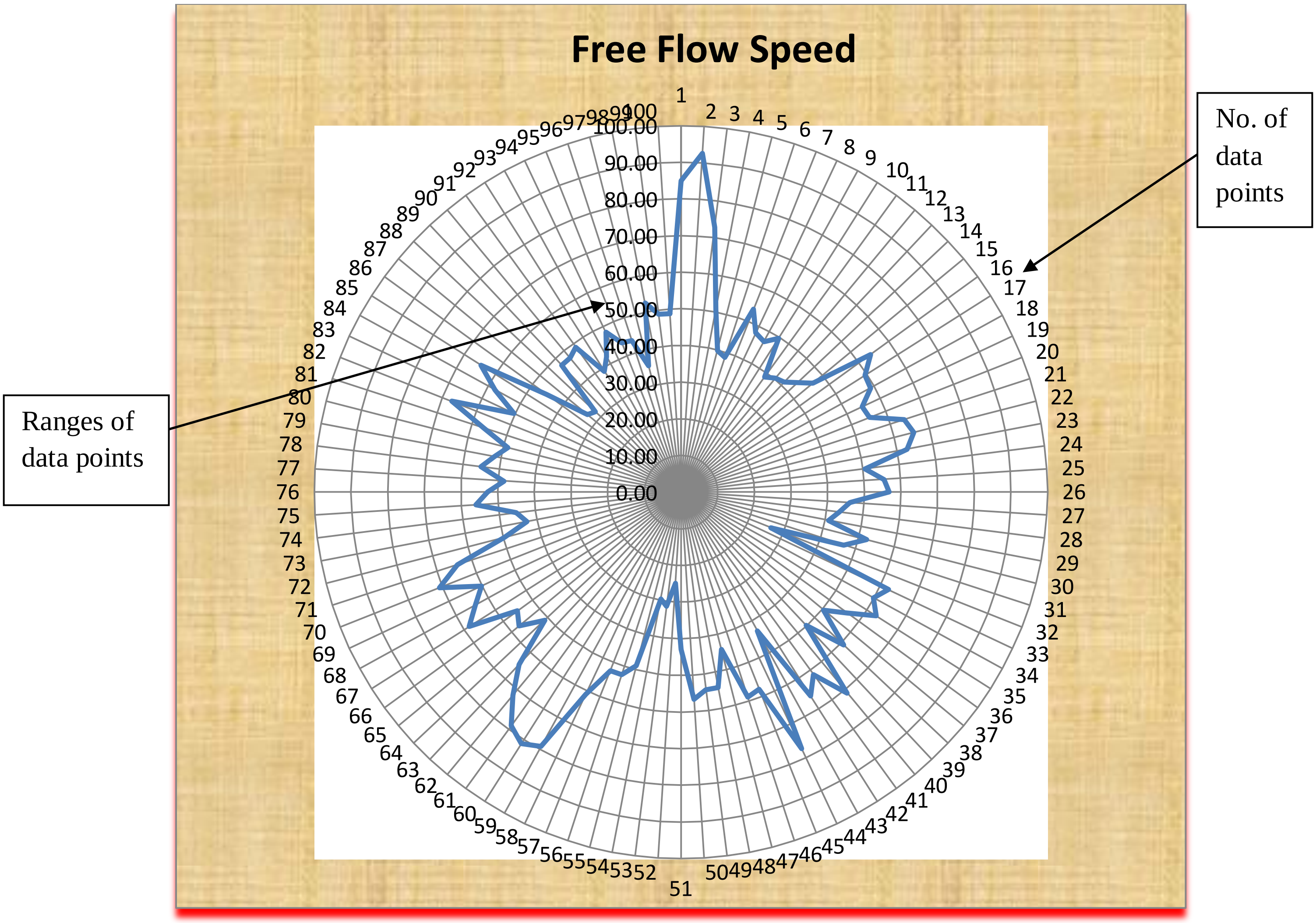


Fig 5.13 Represents of Free Flow Speed (FFS) in Radar diagram

Here in this figure 100 numbers of data points are represented in 100 spokes. The data length of a spoke is proportional to the magnitude of the variable for the data points relative to the maximum magnitude of the variable across all data points. The 100 numbers of data points such as FFS values ranging from [24.94, 26.35, 29.62, 31.38 …………………79.23, 81.30, 85.00, 92.56] are represented in radar diagram. A line is drawn connecting to the 100 numbers of data points for each spoke and finally plots a star like pattern is called radar diagram.

5.4 Analysis of Classification Errors of Free Flow Speed by Adaboost, GP, ML & EM Method:

The urban street classes are classified into suitable number of classes by taking free flow speed data. There are four clustering methods such as (Adaboost, ML, GP & EM) have used for clustering the free flow speed (FFS) data. There are two types of errors have observed i.e. (Test Set Errors & Train Set Errors) in the classification error process. The fraction of mistakes made on training set is known as training error and the fraction of mistakes made on testing set is known as testing error.

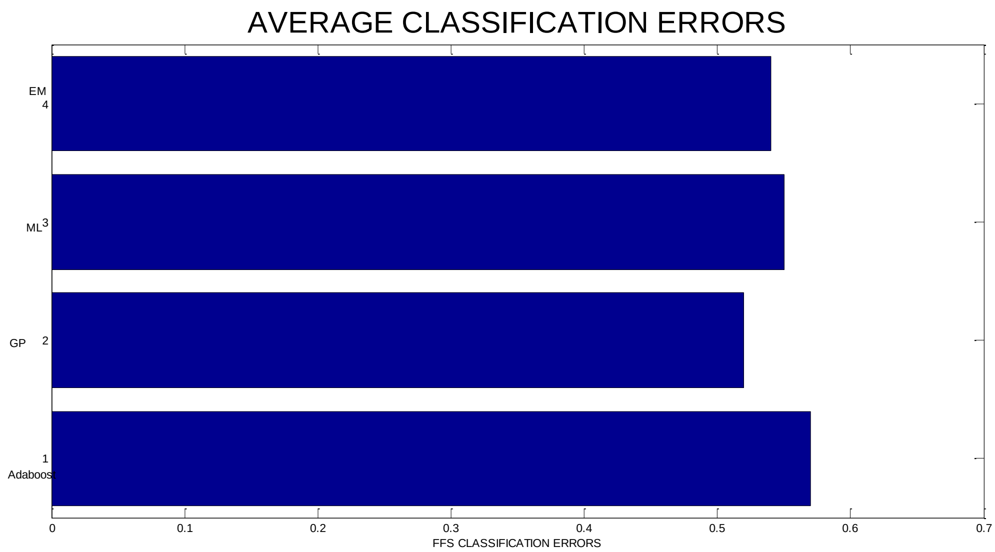


Fig 5.14 Represents the Classification Errors Of Free Flow Speed (FFS) by Four Methods (Adaboost, ML, GP&EM) in Bar diagram

Table-5.5: Represents the Classification Errors of Free Flow Speed (FFS) by four Methods (Adaboost, ML, GP&EM).

Classification Methods	Train Set Errors	Test Set Errors	Total Mean Errors
Adaboost Algorithm	0.67	0.52	0.57
Genetic Programming	**0.55**	**0.49**	**0.52**
Maximum Likelihood Method	0.65	0.46	0.55
Expectation-Maximization Method	0.57	0.51	0.54

The Table 5.5 indicates that the Train Set Errors should be maximum than the Test Set Errors in these four Classification Methods. It is observed that Genetic Programming Method has lowest Total Mean Errors than other three methods (Adaboost, ML & EM). Hence GP is the most suitable method for this study. So GP was selected as the good clustering method in defining LOS criteria in Indian context. Speed ranges obtained from GP clustering is decided to be most relevant in Indian context.

Chapter 6

Summary, Conclusions and Future Scope

6.1 Summary

In this research, there are various limitation to current HCM-2000 methodology for defining LOS criteria in Indian context and various iterative procedure has been made to define LOS criteria for urban streets in Indian context. GPS was used to collect the speed and inventory data and GIS was used handle these data. Applications of GIS and GPS for traffic data collection were reviewed from literature. The concept of urban street classification based on free-flow speeds, function and geometric characteristics of street segments are presented. Also important influencing factors that affect level of service classifications of urban streets are enumerated.

From literature it was found that cluster analysis is the suitable technique that can be applied for the classification of urban streets and level of service categories. Adaboost, Genetic Programming, Maximum Likelihood Method and Expectation Maximization Method were used as a tool to cluster the speed data to classify the road segments into various classes and also to define the speed ranges of the LOS. By using above four algorithms FFS speed were clustered into four different groups corresponding to different urban street classes. Then clustering methods were applied on average travel speeds on street segments of each class of urban street during peak and off peak hours. In the latter case, speeds were classified into six categories for six levels of service; thus speed ranges for level of service categories were defined for Indian conditions. Number of cluster into which the data set should be clustered into is given as prior to the clustering algorithm. To determine the optimum number of clusters various cluster validation parameters were used as not a single parameter is self-sufficient. In this study as a whole 5 validation parameters were used i.e. Homogeneity-Separation Index(HIS), Rand Index(RI), Adjusted Rand Index(ARI), Mirkin Index(MI), Hubert Index(HI).

6.2 Conclusions

The following conclusion are listed below in this research work in defining level of service criteria of roads in urban Indian context.

- Various cluster validation measures, based on their applicability is used to find the optimal number of clusters for Adaboost, Genetic Programming, Maximum Likelihood Method, Expectation Maximization Method. After thorough analysis it was decided to classify urban street into four classes (I-IV) in Indian context. Free flow speed ranges for different urban street classes were found out and for each algorithm the range was found to be different. The speed ranges were lower than that mentioned in HCM-2000. Heterogeneous traffic flow and roads having varying geometric and surrounding environmental characteristics are the major reasons for these lower values in FFSs.

- After determining the FFS ranges of different urban street classes, speed ranges of LOS categories were also found using the four different clustering algorithms. These speed ranges resulted from different clustering algorithm were found to be significantly different from each other. In order to get the most suitable clustering algorithm in defining LOS criteria a thorough study of method of classification errors was carried out. The four clustering method such as adaboost, genetic programming, Maximum-Likelihood and expectation-maximization were tested to find out the classification error rate in the classification error process. The least classification error rate in between the four methods indicates the best clustering methods. The classification errors showed GP to be most suitable clustering algorithm for this study. So GP is selected as the best clustering method in defining LOS criteria in Indian context. Speed ranges obtained from GP clustering is decided to be most relevant in Indian context.

The following LOS criteria for urban streets in Indian context are suggested by GP clustering:

Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	67to 90 km/h	57to 67 km/h	45 to 57 km/h	25 to 45 km/h
Typical FFS	72km/h	61km/h	51km/h	37 km/h
LOS	Average Travel Speed (Km/h)			
A	>71	>52	>45	>39
B	>61-71	>42-52	>34-45	>30-39
C	>49-61	>34-42	>25-34	>24-30
D	>39-49	>28-34	>18-25	>17-24
E	>26-39	>19-28	>11-18	>9-17
F	≤26	≤19	≤11	≤9

- From this research, it was witnessed that the urban street speed ranges valid in Indian context are proportionately lower than that shown in Highway Capacity Manual (HCM 2000). In HCM 2000, the FFS ranges are (90-70) km/hr, (70-55) km/hr, (55-50) km/hr and (55-40) km/hr for class I, II, III, IV respectively. Whereas, by enforcing the Genetic Programming in the FFS data, it resolves that the speed ranges are (67-90) km/hr, (57-67) km/hr, (45-57) km/hr and (25-45) km/hr, which are comparatively lower than shown in HCM 2000.

- FFS range of urban street class IV in particular is very low because of highly heterogeneous traffic flow on urban roads with varying geometry and surrounding environmental features. . For similar reasons it is observed that speed ranges of poor LOS categories such as “E” and “F” under urban street classes III and IV are very low. This implies that the road networks comprise some segments on which traffic moves at stop and go condition.
- From this research output it is found that average speeds of LOS categories (A-F) expressed in terms of percentage of FFS are 90 and above, 75-90,60-75,40-60,30-40,less than equal to 30 respectively. In HCM (2010) these values are shown as 85 and above, 67-85, 50-67, 40-50, 30-40 and less than equal to 30 respectively for LOS categories “A” to “F”.
- Also it is observed from this study that average travel speed expressed in terms of percentage of FFS for LOS category “C” varies from 40 to 60, which is significantly different from that expressed in HCM (2010). The finding implies that large volume of traffic travel at average kind of quality of service on Greater Mumbai road. This result suggests that the road network needs geometric improvements to produce better quality of service.

6.3 Limitation and Future Scope

There are some limitations in this research work and further study can be carried out to overcome these limitations.

- The research is carried out only for the city of Mumbai and this research can be further executed in other cities to determine the Level of Service (LOS) criteria of roads due to heterogeneous of traffic flow, road condition of other cities and driving characteristics.

- The mid-sized vehicle is only used for this research work. Trimble GeoXT GPS receiver fitted on mid-sized vehicles for this research, because it provides consistency, automation, finer levels of resolution and better accuracy in measuring travel time, delay and speed. The study can be further carried out by using other modes of vehicles.
- The user perception should be given consideration in defining LOS criteria of roads in urban Indian context. This study is based on quantitative measure of service, which can be extended for qualitative measurement to develop comprehensive LOS criteria.

References

Arasan, V.T., VedagirI,P. Study of the impact of exclusive bus lane under highly heterogeneous traffic condition. *Public Transport* ,2010, Vol 2(1),pp. 135-15.

Basu, D., Maitra Roy, S., Maitra, B. (2006) "Modeling passenger car equivalency at an urban mid-block using stream speed as measure of equivalency." *European Transport /Trasporti Europei*, 34, pp. 75-87.

Baumgartner, W.E. Level of service: Getting ready for the 21st century. *ITE Journal (Institute of Transportation Engineers),* 1996, Vol. 66 (1), pp. 36-39.

Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P., Silbiger, M.L., Arrington, J.A., Murtagh, R.F. Validity –guided (re) clustering with applications to image segmentation. *IEEE Transactions on Fuzzy Systems*, 1996, Vol 4(2), pp.112-123.

Benz, R.J.; Ogden M.A. 1996. Development and Benefits of Computer Aided Travel Time Data Collection.*Transportation Research Record: Journal of the Transportation Research Board*, 1996: 1-7.

Bhuyan, P.K, Krishna Rao K.V. (2011) "Defining level of service criteria of urban streets in Indian context", *European Transport /Trasporti Europei*, 49, pp. 38-52.

Cao, S.H., Yuan, Z.Z., Zhang, C.Q., Zhao, L. LOS Classification for Urban Rail Transit Passages Based on Passenger Perceptions. *Journal of Transportation Systems Engineering and Information Technology, 2009*, Vol 9 (2), pp. 99-104.

Clark, I. (2008) “Level of Service F: Is it really bad as it gets?” In *IPENZ Transportation Group Conference,* New Plymouth, November.

Cramer, Nichael Lynn (1985), "A representation for the Adaptive Generation of Simple Sequential Programs" in *Proceedings of an International Conference on Genetic Algorithms and the Applications*, Grefenstette, John J. (ed.), Carnegie Mellon University.

Dandan,T., Wei,W., Jian,L., Yang, B. (2007) “Research on Methods of Assessing Pedestrian Level of Service for Sidewalk.*” J Transpn Sys Eng & IT* , 7(5), pp.74−79.

Ernst, Joseph M. et.al., “ Maximum likelihood speed estimation using vechile –induced magnetic signatures”. Proceedings of the 12th international IEEE conference on intelligent transportation systems, St. louis, Mo, USA oct 3-7,2009.

Fang, F.C., Pecheux, K.K. (2009) “Fuzzy data mining approach for quantifying signalized intersection level of services based on user perceptions.” *Journal of Transportation Engineering,* ASCE 135 (6), pp.349-358.

Fang,F.C., Elefteriadou,L., Pecheux,K.K., Pietrucha,M.T. (2003) “Using Fuzzy Clustering of User Perception to Define Levels of Service at Signalized Intersections.” *Journal of Transportation Engineering,* ASCE, 129 (6), pp.657-663.

Flannery, A. ,Rouphail, N., Reinke, D. (2008) “Analysis and Modeling of Automobile Users’ Perceptions of Quality of Service on Urban Streets.” *Transportation Research Record,* 2071, Transportation Research Board, Washington, D.C., pp. 26-34.

Flannery, A., Wochinger, K.., Martin, A. Driver assessment of service quality on urban streets. *Transportation Research Record*, 1920 Transportation Research Board, 2005, Washington, D.C., pp. 25–31.

Forsyth, Richard (1981), BEAGLE A Darwinian Approach to Pattern Recognition Kybernetes, Vol. 10, pp. 159–166

Halkidi, M., Batistakis, Y., Vazirgiannis, M.(2002) Cluster validity checking methods: Part II. ACM *SIGMOD Record Archive* Vol. 31 (3) , pp. 19-27.

Hauchun Tan et. al. “Extracting Auto-correlation feature for licence plate detection based on adaboost”.Proceeding IDEAL 2008,pp.72-79.

Highway Capacity Manual. *Transportation Research Board*, 1950, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 1965, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 1985, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 2000, Washington, D.C.

Hong chen et.al. " The state evaluation of highway traffic flow based on observed data". International conference on transportation engineering 2009, volume1.

IRC 106: Guidelines for Capacity of Urban Roads in Plain Areas. 1990. Indian Road Congress.

Ivana,C., Zvonko,K., Marjana,P. Hybrid approach for urban roads classification based on GPS tracks and road subsegments data. *Promet-Traffic & Transportation*,2011, Vol. 23(4), pp. 289-296.

Jain, A. K.; Dubes, R. C. 1988. *Algorithms for Clustering Data*. Prentice Hall College Div. 304.

Jin Hang, "Extraction of road lines from high resolution stereo aerial imagery based on maximum likelihood segmentation and texture enhancement". Digital image computing techniques and applications, 2009, DICTA'09, 1-3 dec. 2009,pp. 271-276.

Kayabol, Koray et. al. " Un supervised classification of SAR images using EM algorithm". Computational intelligence for multimedia understanding MUSCLE 2011, pp.54-65.

Kikuchi,S., Chakroborty,P. (2007) “Frameworks to Represent the Uncertainty when Determining the Level of Service.” *Transportation Research Record*, 1968, pp. 53-62.

Kita, H., Fujiwara, E. (1995). “Reconsideration on the level of service and a proposed measure.” Proceedings of *15th Annual Meeting of JSTE*, Japanese, pp.25–28.

Kittelson, W.K.., Roess, R.P. Highway capacity analysis after the highway capacity manual 2000. *Transportation Research Record*, 1776, Transportation Research Board, Washington, *D.C.* 2001, pp. 10–16.

Korada, N.K., Kumar N.S.P et.al. “Implementation of Naïve Bayesian classifier and adaboost algorithm using maize –expert system”. International journal of information sciences and techniques (IJIST),vol.2,No.3,May 2012.

Koza, J.R. (1990), *Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems*, Stanford University Computer Science Department technical report STAN-CS-90-1314.

Koza, J.R. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press

Koza, J.R. (1994), *Genetic Programming II: Automatic Discovery of Reusable Programs*, MIT Press

Laxhammar, R. et. al. “Anomally detection for sea surveillance”. Information fusion ,2008 , 11th international conference , june30-july3,2008,pp.1-8.

Lingras, P. 1995. Classifying highways: Hierarchical grouping versus kohonen neural networks, *Journal of Transportation Engineering, ASCE*, DOI: http://dx.doi.org/10.1061/(ASCE)0733947X(1995)121:4(364),121(4): 364-368.

Lingras, P.(2001) “Statistical and Genetic Algorithms Classification of Highways”, Journal of Transportation Engineering, ASCE,127(237-243).

Maitra, B., Sikdar, P.K., Dhingra, S.L. Modelling congestion on urban roads and assessing level of service. *Journal of Transportation Engineering*, 1999, Vol 125 (6), ASCE, pp. 508-514.

Maitra,B., Sikdar, P. K., Dhingra, S. L. (2004) "Modeling of Congestion: A Tool for Urban Traffic Management in Developing Countries", *European Transport / Trasporti Europei*, 27, pp. 45-56.

Marwah, B.R., Singh, B. Level of service classification for urban heterogeneous traffic: A case study of Kanpur metropolis. In Proc., *Fourth International Symposium on Highway Capacity,* Hawaii, June-July, 2000, pp. 271-286.

Meila, M.,(2007). Compairing Clusters- An Information based distance. Journal of Multivariate Analysis, 98(5), pp.873-895.

Pattnaik, S.B., Ramesh Kumar, K. (1996). “Level of service of urban roads based on users' perception.” Civil Engineering Systems, 14, pp.87-110.

Pecheux, K.K., Pietrucha, M.T., Jovanis, P.P. (2000) “User perception of level of service at signalized intersections: Methodological issues.” *79th Annual Meeting of TRB,* Transportation Research Board, Washington, D.C., pp.322-335.

Pedro G.espejo et al.(2010), “ Web Usage Mining for predicting final marks of students that use moodle courses”, Wiley Periodicals, CAE 09-037.R1(20456)pp.1-11._

Prasses, E.S.,Roess, R.P., Mcshane,W.R.(1996),Cluster analysis as tool in traffic engineering”.Transportation Research Record, 1551, TRB, National Research Council, Washington, D.C., pp-39-48.

Quiroga, C., Bullock, D. (1998). “Travel time studies with global positioning and geographic information systems: An integrated methodology.” *Transportation Research Part C*, 6(1–2), pp.101–127.

Shamir R.,Sharan R. et al; Cluster Graph Modification Problems, In proceedings of the 27th international workshop Graph-theoretic concepts in computer science(WG),2002.

Shao, M.,Sun, L. (2010) "United evaluation model of traffic operation level for different types of urban road" *Journal of Tongji University,* 38 (11) ,pp.1593-1598.

Shivakumar B.R. et. al., "Fuzzy logic based RS image classification using maximum likelihood and mahalanobis distance classifiers". International journal of current engineering and technology , ISSN 2277-4106. Vol(3). No. 2 (2013).

Shouhua,C., Zhenzhou, Y., Chiqing, Z., Li,Z. (2009) "LOS Classification for Urban Rail Transit Passages Based on Passenger Perceptions." *J Transpn Sys Eng & IT,* 9(2), pp.99-104.

Spring, G.S. (1999) ''Integration of safety and the highway capacity manual.'' *Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity,* Transportation Research Board, Washington, D.C., pp. 63–72.

Tan, D., Wang, W., Lu, J., Bian, Y. Research on Methods of Assessing Pedestrian Level of Service for Sidewalk. *Journal of Transportation Systems Engineering and Information Technology,* 2007, Vol 7 (5),pp. 74-79.

Tsai, Andy et.al. "An EM algorithm for shape classification based on level sets". Medical image analysis 9(2005), pp. 491-502.

Turner, et.al. *Travel time data collection handbook*, Texas Transportation Institute, The Texas A&M Univ. System, College Station, Texas, 1998.

Van Dongen, Stijn: Performance Criteria for Graph Clustering and Markov Cluster Experiments. Technical Report INS-R0012,centrum voor wiskunde en informatica,2000.

Weinert Antje, "Estimation of critical gaps and follow-up times at rural un-signalized intersections in germany". Transportation Research circular E-C018: 4th international symposium on highway capacity.

Xiaong Weng Xiao "Analysis on capacity of road traffic based maximum likelihood methods". Information and automation, 2009, ICIA' 09. 22-24, June, 2009. pp.540-543.

Yuan-cheng Xie et. al. "Using Boosting and clustering to prune bagging and detect noisy data. Pattern recognition, 2009, Chinese conference, pp.- 1-5.

Zhaufeng, H.E. et.al. "Topology modelling for adaboost cascade based object detection".Elsevier, pattern recognition letters 31(2010), pp.912-91

.

Appendix-I

The table illustrates FFS and average travel speed of 19 segments belongs to corridor-5.

Table A-I. FFS and Average travel speed during peak and off peak hours of Corridor-5

Corridor-5					
Segment No.	Average Free-Flow Speed (km/hr)	Duration and Direction of Travel			
		M-E-W	M-W-E	E-E-W	E-W-E
		Average Travel Speed (km/hr)			
1	67.28	59.28	73.78	57.44	69.13
2	50.39	43.31	35.82	33.76	14.42
3	57.86	47.15	44.26	48.19	43.18
4	64.64	66.21	46.50	59.80	52.88
5	45.26	17.49	43.50	29.23	33.52
6	33.14	12.82	15.12	12.72	14.95
7	32.12	14.19	15.07	11.86	11.84
8	47.52	17.40	17.62	12.29	16.32
9	47.53	14.85	5.86	13.47	14.97
10	48,92	15.59	14.74	20.94	19.09
11	39.03	15.97	5.41	17.88	15.08
12	42.17	33.38	13.80	6.45	18.13
13	48.26	9.55	15.42	6.58	23.33
14	43.79	14.18	22.80	20.19	10.13
15	43.56	11.15	12.07	19.69	20.44
16	35.53	13.78	14.91	12.56	14.15
17	52.49	9.58	20.49	13.20	14.96
18	48.96	8.20	13.11	6.97	14.90
19	48.86	17.09	16.21	22.05	12.18

Note:

M-E-W= Morning -East-West

M-W-E= Morning -West-East

E-E-W= Evening- East -West

E-W-E= Evening -West -East

List of Publications

Journal

1. Patnaik, A.K., Bhuyan, P.K. (2013) "Application of GP for delimitating LOS measures for heterogeneous traffic flow". *IATSS research, Elsevier, (submitted).*
2. Patnaik, A.K., Das, A.K, Dehury, A.N., Bhuyan, P.K., Chattaraj, U., Panda, M.,(2013) "Adaboost clustering in defining LOS criteria of Mumbai city". *IJEI-e-ISSN:-2278-7461, P-ISSN:-2319-6491, Volume 2, Issue-8, pp. 45-55.*
3. Das, A.K., Patnaik, A.K., Dehury, A.N., Bhuyan, P.K., Chattaraj, U., Panda, M., (2013) "Defining Level of Service criteria of urban streets using Neural Gas clustering". *IOSRJEN, (Accepted).*
4. Dehury, A.N., Patnaik, A.K., Das, A.K., Chattaraj, U., Bhuyan, P.K., Panda, M., (2013) "Accident analysis and modelling on NH-55". *IJEI, ISSN:2278-7461, ISBN-2319-6491, Volume 2, Issue-7,pp.80-85.*
5. Dehury, A.N., Patnaik, A.K., Das, A.K., Chattaraj, U., Bhuyan, P.K., Panda, M., (2013) "Black spot analysis on national highways". *IJERA, ISSN:-2248-9622, vol.(3), issue 3,pp.402-418.*
6. Dehury, A.N., Patnaik, A.K., Das, A.K., Chattaraj, U., Bhuyan, P.K., Panda, M.,.(2013) "Accident analysis on two lane highways". *IJETAE, ISSN:-2250-2459, Volume 2,issue-6, (Submitted).*
7. Patnaik, A.K., Bhuyan, P.K. (2013) "A comparative study between EM & ML clustering Algorithm for delineating LOS measures in Indian city". *Road & Transport research journal, Australia, (Under Preparation).*